CALMAR 2: A NEW VERSION OF THE CALMAR CALIBRATION ADJUSTMENT PROGRAM

Olivier Sautory[1]

ABSTRACT

Calmar 2 is the new version of the Calmar calibration adjustment program. It contains two major developments.

When survey data are collected at different levels (e.g., households and individuals), simultaneous calibration of the samples helps maintain consistency in the statistics produced from the samples.

Where there is total non-response, generalized calibration makes it possible to rewrite the calibration equations with two sets of variables: the actual calibration variables and the non-response explanatory variables. This corrects for non-response even when the variables that explain it are unknown for the sample non-respondents.

KEYWORDS: Calibration; Generalized Calibration; Non-Response; Simultaneous Calibration.

1. THE CALMAR MACROS

1.1 Background

Calmar is a SAS macro program that implements the calibration methods developed by Deville and Särndal (1992). The program adjusts samples, through reweighting of individuals, using auxiliary information available from a number of variables referred to as calibration variables. The weights produced by this method are used to calibrate the sample on known population totals in the case of quantitative variables and on known category frequencies in the case of qualitative variables.

Calmar is an acronym for CALibration on MARgins, an adjustment technique which adjusts the margins (estimated from a sample) of a contingency table of two or more qualitative variables to the known population margins. However, the program is more general than mere "calibration on margins," since it also calibrates on the totals of quantitative variables.

Calmar was developed in 1990 at France's Institut National de la Statistique et des Études Économiques (INSEE), where it is used regularly to adjust survey data. It is also used by many other statistics agencies in France and other countries.

The new version, Calmar 2, developed in 2003, offers the user new resources for performing calibrations and implements the generalized calibration method of handling non-response proposed by Deville (1998).

Calmar can be downloaded from INSEE's Web site (www.insee.fr), and Calmar 2 will also be available on the site sometime in 2006.

1.2 Calmar's calibration methods

It is worth restating the principle underlying the calibration methods implemented by Calmar (see also Deville *et al*., 1993).

Consider a population U of individuals, from which a probabilistic sample *s* has been selected. Let ***Y*** be a variable of interest, for which we want to estimate the total in the population : $Y = \sum_{k \in U} y_k$.

[1] Olivier Sautory, Cepe-Insee, 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France, sautory@ensae.fr.
This text has been first published in the Statistics Canada International Symposium Series - Proceedings, 2003.

The usual estimator of Y is the Horvitz-Thompson estimator :

$$\hat{Y}_{HT} = \sum_{k \in s} \frac{1}{\pi_k} y_k = \sum_{k \in s} d_k y_k \ .$$

Assume that we know the population totals for J auxiliary variables[2] *X*1… *X j*… *X J* available in the sample:

$$X_j = \sum_{k \in U} x_{jk}$$

We will look for new "calibration weights" *wk* that are as close as possible (as determined by a certain distance function) to the initial weights *dk* (these are usually the "sampling weights," equal to the inverses of the probabilities of inclusion π_k). These *wk* are calibrated on the totals of the *X j* variables; in other words, they verify the **calibration equations**:

$$\forall j = 1...J \quad \sum_{k \in s} w_k x_{jk} = X_j \qquad (1)$$

The solution to this problem is given by $w_k = d_k F(x_k' \lambda)$ where $x_k' = (x_{1k}...x_{Jk})$, λ is a vector of J Lagrange multipliers associated with the constraints (1), and F is a function – the calibration function – whose terms depend on the distance function that is used.

Vector λ is determined by the solution to the non-linear system of J equations in J unknowns resulting from the calibration equations:

$$\sum_{k \in s} d_k F(x_k' \ \lambda) x_k = X$$

The estimator of the total for a variable of interest will be the "calibrated" estimator $\hat{Y}_w = \sum_{k \in s} w_k y_k$.

The original version of Calmar offered four calibration methods, corresponding to four different distance functions. These methods are characterized by the form of function F:

- the *linear* method: the calibrated estimator is the generalized regression estimator:

$$\hat{Y}_{reg} = \hat{Y}_{HT} + \left(X - \hat{X}_{HT}\right)\hat{B}_s, \quad \text{where} \quad \hat{B}_s = \left(\sum_{k \in s} d_k \ x_k \ x_k'\right)^{-1} \left(\sum_{k \in s} d_k \ x_k \ y_k\right)$$

- the *exponential* method: where all the calibration variables are qualitative, this is the raking ratio method
- (Deming and Stephan, 1940) ;
- the *logit* method: this method provides lower limits L and upper limits U on the weight ratios *wk*/*dk* ;
- the *truncated linear* method, very similar to the logit method.

The last two methods are used to control the range of the distribution of weight ratios. The logit method is used more often because it avoids excessively large weights, which can compromise the robustness of the estimates, and excessively small or even negative weights, which can be produced by the linear method.

[2] Quantitative variables or indicators associated with the response categories of qualitative variables.

Precision

All of the Y_w calibrated estimators have the same precision (asymptotically), regardless of the method used: the approximate variance of *Yw* is therefore equal to that of the regression estimator $\hat{Y}_{reg}$:

$$AV\left(\hat{Y}_w\right) = \sum_U \sum \Delta_{k\ell} \left(d_k E_k\right)\left(d_\ell E_\ell\right) U$$

where $E_k = y_k - x_k' B$, with $B = \left(\sum_{k \in U} x_k x_k'\right)^{-1} \left(\sum_{k \in U} x_k y_k\right)$, and *Ek* the residual of the regression of ***Y*** on the *X j* in the population U.

This variance is especially small if the variable of interest ***Y*** and calibration variables *X*1… *X j*… *X J* are strongly correlated.

A variance estimator is given by :

$$\hat{V}\left(\hat{Y}_w\right) = \sum_s \sum \frac{\Delta_{k\ell}}{\pi_{k\ell}} \left(d_k e_k\right)\left(d_\ell e_\ell\right),$$

where $e_k = y_k - x_k' B_s$, avec $B_s = \left(\sum_{k \in s} d_k x_k x_k'\right)^{-1} \left(\sum_{k \in s} d_k x_k y_k\right)$, and *ek* is the residual of the regression (weighted by the *dk*) of ***Y*** on the *X j* in sample *s*.

1.3 What's new in Calmar 2

In addition to the four calibration methods mentioned above, Calmar 2 (Le Guennec and Sautory, 2003) offers the following :

- simultaneous calibration for different levels in a survey;
- adjustment for total non-response using generalized calibration.

These two features will be described in detail below.

Calmar 2 offers a solution to the problem of collinearity between calibration variables: it uses generalized inverse matrices to compute the calibration weights. The original version of Calmar produced an error message in such cases.

Calmar 2 also offers a new distance function, the *generalized hyperbolic sine* function, which depends on a parameter α. Like the exponential method, this method always yields positive weights, but the distribution of weights at the high end is narrower. In addition, the method reduces the range of the weight distribution, as do the logit and truncated linear methods, but it does so with only one parameter, α (Roy *et al.*, 2001).

Finally, the program is more user-friendly, especially in two respects:

- users can enter qualitative calibration variables without prior recoding to obtain sequential response categories ;
- users have the option of entering parameters interactively using capture screens that guide them in their choices.

2. SIMULTANEOUS CALIBRATIONS

2.1 The problem

In some surveys, data are collected at different levels:

- INSEE's continuing survey of household living conditions includes questions about the household (type of dwelling, number of persons, occupation of the head of the household, etc.), each member of the household (sex, age, occupation, etc.) and usually a specific set of questions for an individual selected at random from the "eligible" members of the household (often those aged 15 and over), referred to as the "Kish individual";
- the French industry ministry's annual business survey contains questions on each firm's overall activities and a section on each of its establishments.

When the survey data are adjusted, either independent calibrations can be performed for the various levels or simultaneous ("combined") calibrations can be carried out. Simultaneous calibration produces the same weights for all members of a household provided they were all surveyed and ensures consistency in the statistics obtained from the various data files. For example, when independent calibrations are performed on the sample of households and on the sample of household members, the number of one-person households estimated from the former sample cannot be expected to match the number of persons belonging to one-person households estimated from the latter sample.

2.2 The method

More generally, the situations described above relate to surveys that involve cluster sampling or multi-stage sampling, where there is auxiliary information about the clusters (or primary units) and the secondary units, and where the survey's variables of interest concern both the clusters (or PUs) and the SUs.

The simultaneous calibration method was proposed by Sautory (1996). It is more general than the method proposed by Lemaître and Dufour (1987). It consists in performing a single calibration at the PU level. Estimates of the totals for the calibration variables defined at the SU level are computed and then used in the PU calibration, which includes both PU and SU variables.

Thus, if $\boldsymbol{X}$ is a calibration variable for the SUs, the estimate $\hat{X}_m = \sum_{k \in m} \pi_{k/m} x_k$ is calculated for each PU *m*, where $\pi_{k/m}$ denotes the probability of inclusion of SU *k* when PU *m* has been selected. Hence, the calibration equation for variable $\boldsymbol{X}$ can be written $\sum_{m \in s_M} w_m \hat{X}_m = X$, where s_M denotes the PU sample.

2.3 An example

Suppose we have a survey in which a sample of households s_M was selected and some data on the sample households were collected. All members of the selected households were surveyed, forming a sample sI. In addition, an individual k_m (referred to as the Kish individual) was chosen in each selected household *m* by simple random sampling without replacement among the e_m eligible members of the household (e.g., individuals aged 15 and over) and surveyed with a specific questionnaire.

Note that:

xm is the vector of known auxiliary variables for each household *m* in household sample s_M,

$X = \sum_{m \in U_M} x_m$ is the vector of the totals for these variables, which totals are known for the population of households U_M

$z_{m,i}$ is the vector of known auxiliary variables for each individual *i* in household *m*,

$Z = \sum_{i \in U_I} z_i$ is the vector of the totals for these variables whose totals are known for the population of individuals U_I,

v_{k_m} is the vector of known auxiliary variables for each Kish individual k_m in household *m* and

$V = \sum_{i \in U_I^e} v_i$ is the vector of the totals for these variables whose totals are known for the population of eligible individuals U_I^e.

The probabilities of inclusion of households *m* are denoted π_m, and we let $d_m = 1/\pi_m$. The probabilities of inclusion of individuals (*m,i*) when household *m* has been selected are 1. The probability of inclusion of Kish individual k_m when household *m* has been selected is $1/e_m$.

The method involves performing a single calibration at the household level, calculating for each household *m* the totals of the calibration variables for individuals $Z_m = \sum_{(m,i) \in men_m} z_{m,i}$, and the estimated totals of the calibration variables for Kish individuals $\hat{V}_m = e_m v_{k_m}$.

The calibration variables vector for household *m* becomes $(x_m, Z_m, \hat{V}_m)$, and the totals vector (X, Z, V). The calibration equations are written as follows:

$$\sum_{m \in s_M} d_m F(x'_m \lambda + Z'_m \mu + \hat{V}'_m \gamma)(x_m, Z_m, \hat{V}_m) = (X, Z, V)$$

λ, μ, γ denote components of the Lagrange multipliers vector.

The solutions $w_m = d_m F(x'_m \lambda + Z'_m \mu + \hat{V}'_m \gamma)$ of these equations are the new household weights. Thus, the weight $w_{m,i}$ assigned to individual *i* of household *m* in the sample of individuals is equal to the weight w_m of household *m*.

The weight w_{k_m} assigned to the Kish individual of household *m* is equal to $e_m w_m$. It can be verified that with these weights, the various samples are correctly calibrated on totals *X*, *Z* and *V*:

$$\sum_{i \in s_I} w_{m,i} z_{m,i} = \sum_{m \in s_M} w_m \left(\sum_{(m,i) \in men_m} z_{m,i} \right) = \sum_{m \in s_M} w_m Z_m = Z$$

$$\sum_{k_m \in s_K} w_{k_m} v_{k_m} = \sum_{k_m \in s_K} w_m e_m v_{k_m} = \sum_{k_m \in s_K} w_m \hat{V}_m = V$$

This method could be used with Calmar (see Caron and Sautory, 2004), but some SAS programming would be required. Calmar 2 performs all the operations necessary to reduce the process to a single calibration. The user must provide the entry tables for the various levels and the totals for the calibration variables. Estevao and Särndal (2003) compare several calibration methods for two-stage sample designs, including the method described below.

3. GENERALIZED CALIBRATION

3.1 The underlying principle

While calibration is usually presented using functions of distance between weights, Deville (2002, for example) states the calibration equations directly, with calibration functions defined in a very general form: $F_k : \lambda \in \mathbb{R}^J \to F_k(\lambda) \in \mathbb{R}$ □, with $F_k(0) = 1$, where λ is a vector of *J* adjustment parameters.

The generalized calibration equations are written $\sum_{k \in s} d_k \, F_k(\lambda) x_k = X$, where x_k, as previously, denotes the vector of the *J* calibration variables. Solving this system for λ yields the new weights $w_k = d_k \, F_k(\lambda)$ □.

Basic result

Let □□□□ grad $F_k(0) = z_k$, vectors that will be referred to as "instruments" (see below). We can show that calibrated estimators based on the same instruments and the same calibration variables are all asymptotically equivalent.

We can rewrite the calibration equations □□□□ $X = \sum_{k \in s} d_k \left(1 + z_k' \lambda + O\|\lambda\|^2\right) x_k$ □.

This yields $X - \hat{X}_{HT} = \left[\sum_{k \in s} d_k \, x_k \, z_k'\right] \lambda + \sum_{k \in s} d_k \, x_k \, O\|\lambda\|^2$, or

$$\lambda_s = (T_{szx}')^{-1} \left(X - \hat{X}_{HT}\right) + O\left\|X - \hat{X}_{HT}\right\|^2,$$

if we let $T_{szx} = \sum_{k \in s} d_k \, z_k \, x_k'$ which is assumed to be of full rank.

A calibrated estimator $\hat{Y}_w = \sum_{k \in s} w_k \, y_k$ is therefore asymptotically equivalent to

$$\sum_{k \in s} d_k \left[1 + z_k' \left(T_{szx}'\right)^{-1} \left(X - \hat{X}_{HT}\right)\right] y_k = \hat{Y}_{HT} + \left(X - \hat{X}_{HT}\right)' T_{szx}^{-1} \sum_{k \in s} d_k \, z_k \, y_k = \hat{Y}_{HT} + \left(X - \hat{X}_{HT}\right)' \hat{B}_{szx} = \hat{Y}_{regi}$$

$\hat{B}_{szx}$ verifies $\sum_{k \in s} d_k \, z_k \, y_k = \left(\sum_{k \in s} d_k \, z_k \, x_k'\right) \hat{B}_{szx}$ □: it is the vector of the coefficients of the instrumental variable regression (weighted by the d_k) of ***Y*** on the $\boldsymbol{X_1} \ldots \boldsymbol{X_j} \ldots \boldsymbol{X_J}$ variables ; the variables that make up the z_k vectors are the "instruments" (for example, see Fuller, 1987). By analogy with the generalized regression estimator, the estimator $\hat{Y}_{regi}$ is referred to as the instrumental variable regression estimator.

3.2 Standard form of the calibration functions

In practice, calibration functions F_k are generally of the form $F_k(\lambda) = F(z_k' \lambda)$ □□, where z_k is a vector of *J* variables $\boldsymbol{Z_j}$ known for sample *s*, and *F* is a function of $\mathbb{R}$ in $\mathbb{R}$ such that $F(0) = 1$ and $F'(0) = 1$ (and hence □□□□ grad $F_k(0) = z_k$).

The calibration equations are $\sum_{k \in s} d_k \, F(z_k' \lambda) x_k = X$.

When *F* is a linear function *F*□*zk*′□□□□□1□□*zk*′□□□, the calibrated estimator is the instrumental variable regression

estimator *Y*ˆ*reg i* , since we have □□□□□□□□*s* □□*Ts*′*zx* −1 *X* −□*X*ˆ *HT* .

3.3 Precision

Through proofs similar to the ones used in conventional calibration, we obtain the following results.

The approximate variance of the calibrated estimator can be written □□□□□□□□□□□ $\mathrm{AV}\left(\hat{\mathrm{Y}}_{\mathrm{w}}\right)=\sum_{\mathrm{U}}\sum\Delta_{\mathrm{k}\ell}\left(\mathrm{d}_{\mathrm{k}}\,\mathrm{E}_{\mathrm{k}}\right)\left(\mathrm{d}_{\ell}\,\mathrm{E}_{\ell}\right)$, where

$\mathrm{E}_{\mathrm{k}}=\mathrm{y}_{\mathrm{k}}-\mathrm{x}_{\mathrm{k}}'\,\mathrm{B}_{\mathrm{zx}}$, avec $\mathrm{B}_{\mathrm{zx}}=\left(\sum_{\mathrm{k}\in\mathrm{U}}\mathrm{z}_{\mathrm{k}}\,\mathrm{x}_{\mathrm{k}}'\right)^{-1}\left(\sum_{\mathrm{k}\in\mathrm{U}}\mathrm{z}_{\mathrm{k}}\,\mathrm{y}_{\mathrm{k}}\right)$ is the residual of the regression of $\boldsymbol{Y}$ on the $\boldsymbol{X}_j$

in U, with instrumental variables $\boldsymbol{Z}_j$.

A variance estimator is given by $\hat{\mathrm{V}}\left(\hat{\mathrm{Y}}_{\mathrm{w}}\right)=\sum_{\mathrm{s}}\sum\frac{\Delta_{\mathrm{k}\ell}}{\pi_{\mathrm{k}\ell}}\left(\mathrm{d}_{\mathrm{k}}\,\mathrm{e}_{\mathrm{k}}\right)\left(\mathrm{d}_{\ell}\,\mathrm{e}_{\ell}\right)$,

where $\mathrm{e}_{\mathrm{k}}=\mathrm{y}_{\mathrm{k}}-\mathrm{x}_{\mathrm{k}}'\,\mathrm{B}_{\mathrm{szx}}$, avec $\mathrm{B}_{\mathrm{szx}}=\left(\sum_{\mathrm{k}\in\mathrm{s}}\mathrm{d}_{\mathrm{k}}\,\mathrm{z}_{\mathrm{k}}\,\mathrm{x}_{\mathrm{k}}'\right)^{-1}\left(\sum_{\mathrm{k}\in\mathrm{s}}\mathrm{d}_{\mathrm{k}}\,\mathrm{z}_{\mathrm{k}}\,\mathrm{y}_{\mathrm{k}}\right)$, is the residual of the regression (weighted

by the *dk* of $\boldsymbol{Y}$ on the $\boldsymbol{j}\,\boldsymbol{X}$ in sample *s*, with instrumental variables $\boldsymbol{j}\,\boldsymbol{Z}$).

4. CALIBRATION IN THE CASE OF TOTAL NON-RESPONSE

4.1 Standard methods of correcting for total non-response

Total non-response is usually accounted for by reweighting the respondent units. Reweighting techniques are based on models of the response mechanism. This mechanism is similar to random selection of a sample *r* (of size *nr*) from sample *s*. This selection can be viewed as a supplementary phase added to the original sample design, defined by a pseudo sample design denoted *q(r|s)*. Associated with this design are the individual response probabilities $\mathrm{p}_{\mathrm{k}}=\mathrm{P}\left(\mathrm{k}\in\mathrm{r}/\mathrm{k}\in\mathrm{s}\right)$.

If these probabilities were known, the total *Y* for a variable of interest would be estimated without bias by

$\hat{\mathrm{Y}}_{\mathrm{exp}}=\sum_{\mathrm{k}\in\mathrm{r}}\frac{1}{\pi_{\mathrm{k}}\,\mathrm{p}_{\mathrm{k}}}\mathrm{y}_{\mathrm{k}}$, known as the expansion estimator. In fact, though, the design *q(r|s)* and therefore the

probabilities *pk* are unknown. They must therefore be estimated, substituting a model for the response mechanism and using an estimation method (maximum likelihood, moments, etc.).

A logical choice is the Poisson model: $q(r|s=\prod_{\mathrm{k}\in\mathrm{r}}\mathrm{p}_{\mathrm{k}}\prod_{\mathrm{k}\in\mathrm{s}\backslash\mathrm{r}}(1-\mathrm{p}_{\mathrm{k}})$. To fully specify this model, we must provide the form of the probabilities *pk* . Three conventional models of the non-response mechanism are described below.

Uniform response model

We assume that each individual has the same probability of response: $\mathrm{p}_{\mathrm{k}}=\mathrm{p}\ \forall\,\mathrm{k}\in\mathrm{U}$. The maximum likelihood method produces the estimate $\hat{\mathrm{p}}=\frac{\mathrm{n}_{\mathrm{r}}}{\mathrm{n}}=$ the observed response rate.

Homogeneous response groups

Population *U* is split into *H* groups that are assumed homogeneous with respect to non-response. All individuals in group *h* have the same probability of response, denoted p_h. The maximum likelihood method produces the estimates $\hat{p}_h = \frac{n_{rh}}{n_h}$, where n_h (resp. n_{rh}), is the number of individuals in group *h* who are in sample *s* (resp. sample *r*). $\hat{p}_h$ is therefore the observed response rate in group *h*.

Generalized linear model

The probability of response is a function of a vector z_k of non-response explanatory variables $\boldsymbol{Z}_j$ and an unknown parameter β : $p_k = 1/H(z_k'\beta)$, where *H* is a function defined on $\mathbb{R}$ with values in $[1,+\infty[$ (in principle). To estimate β and therefore the p_k, the $\boldsymbol{Z}_j$ variables must be known for both respondents and non-respondents.

It is possible to use an even more general model of the form $p_k = 1/H_k(\beta)$, where β is a vector of *J* adjustment parameters and H_k is a function dependent on individual *k*.

We will now examine various calibration strategies for cases where there is total non-response.

4.2 Calibration after correcting for non-response

Suppose we have corrected for total non-response, for example with one of the methods described above. Thus, we can perform a conventional calibration starting with the weights corrected for non-response $d_k^* = \frac{d_k}{\hat{p}_k}$. The calibration equations are written $\sum_{k\in r} d_k^* F^*(x_k'\lambda)\, x_k = X$, , where F^* is one of the usual calibration functions.

4.3 Direct conventional calibration

Another strategy is to perform a calibration directly, without prior correction for non-response. The calibration equations are $\sum_{k\in r} d_k F(x_k'\lambda)\, x_k = X$.

If one of the calibration variables is the constant variable equal to 1, or at least a qualitative variable, the d_k can be multiplied by a constant, with no effect on the $w_k = d_k F(x_k'\lambda)$. Consequently, the calibration equations can be rewritten $\sum_{k\in r} d_k \frac{1}{n_r/n} F(x_k'\lambda) x_k = X$, which shows that this strategy is equivalent to the previous one with a non-response correction using a uniform response model.

Dupont (1996) compared the two strategies on the basis of theoretical considerations and simulations. His study led to the following findings:

- If the non-response correction is performed by a generalized linear model, where the *H* function is one of the usual calibration functions *F*, and if the calibration variables $\boldsymbol{X}_j$ contain the non-response explanatory variables $\boldsymbol{Z}_j$, then the two strategies produce very similar results.

- Furthermore, if the calibration variables $\boldsymbol{X}_j$ are identical with the non-response explanatory variables $\boldsymbol{Z}_j$, the following two strategies are equivalent:

- performing a non-response correction using a generalized linear model with the exponential function as the *H* function, then performing a calibration using the corrected weights with the exponential function as the calibration function *F** ;

- performing a direct calibration using the initial weights with the exponential function as the calibration function *F*.

The same is true if we perform a non-response correction using a homogeneous response group model and then a post-stratification, where the groups and the post-strata are identical. This is equivalent to performing a direct "formal" post-stratification on the respondent sample.

The advantage of direct calibration is that it does not require explicit modelling of the response mechanism. Lundström and Särndal (1999) also studied the properties of direct calibration and in particular, they proposed variance estimators that take sampling variance and non-response variance into account.

4.4 Direct generalized calibration

Let's start with a system of calibration equations on the respondent sample, of the form $\sum_{k\in r} d_k\, H_k(\beta)\, x_k = X$.

These equations can be interpreted as indicated below.

Let there be a response model of the form $p_k = \frac{1}{H_k(\beta_0)}$, where β_0 denotes the actual value of the model's

parameter. The calibration equations can be rewritten as follows, where $\hat{\beta}$ denotes the solution to the system :

$$X = \sum_{k\in r} d_k\, H_k(\beta_0)\frac{H_k(\hat{\beta})}{H_k(\beta_0)}x_k = \sum_{k\in r} d_k\, H_k(\beta_0)\frac{H_k(\beta_0+\lambda)}{H_k(\beta_0)}x_k = \sum_{k\in r} d_k\,\frac{1}{p_k}F_k(\lambda)\,x_k$$

$$\text{with}\quad \hat{\beta} = \beta_0 + \lambda \quad et \quad F_k(\lambda) = \frac{H_k(\beta_0+\lambda)}{H_k(\beta_0)}$$

Hence, these equations take the form of generalized calibration equations, where the initial weights are the d_k / p_k, i.e. the sampling weights corrected for non-response and the functions F_k, which verify $F_k(0)=1$, are the calibration functions. The instruments are $z_k^* = \text{grad}\, F_k(0) = \frac{1}{H_k(\beta_0)}\text{grad}\, H_k(\beta_0)$.
Solving this system is equivalent to performing a non-response correction and a generalized calibration at the same time.

The results in section 3.3 can be used to calculate the precision of the estimators calibrated by this method.

The approximate variance $AV(\hat{Y}_w)$ uses the residuals of the instrumental variable regression in the population $E_k = y_k - x_k' B_{z^*x}$.

The variance estimator $\hat{V}(\hat{Y}_w)$ uses the residuals of the instrumental variable regression in the respondent sample, weighted by the $d_k H_k(\beta_0)$: $e_k = y_k - x_k' B_{rz^*x0}$, where $\sum_{k\in r} d_k\, H_k(\beta_0)\, z_k^* \left(y_k - x_k'\, B_{rz^*x0}\right) = 0$. B_{rz^*x0} is the estimator of B_{z^*x}, which would be calculated if the response probabilities $H_k^{-1}(\beta_0)$ were known. These probabilities are unknown because of β_0; they are estimated by replacing β_0 with $\hat{\beta}$. The residuals become $e_k = y_k - x_k' \hat{B}_{rz^*x}$, where

$\sum_{k \in r} d_k \, H_k\left(\hat{\beta}\right) z_k^* \left(y_k - x_k' \, \hat{B}_{rz^*x}\right) = 0$, which is an instrumental variable regression in sample *r*, weighted by the calibration weights $w_k = d_k \, H_k\left(\hat{\beta}\right)$.

Note : The estimated variance $\hat{V}\left(\hat{Y}_w\right)$ is written in the form $Q_1(e_k) + Q_2(e_k)$, where the quadratic form $Q_1(e_k)$ denotes the phase 1 (selection of sample *s*) variance estimate and $Q_2(e_k)$ denotes the phase 2 ("selection" of sample *r*) variance estimate.

Case of a generalized linear model

In practice, the functions $H_k(\beta)$ are of the form $H(z_k' \beta)$, where z_k is a vector of non-response explanatory variables $\boldsymbol{Z}_j$. The calibration equations are

$$\boxed{\sum_{k \in r} d_k \, H(z_k' \, \beta) x_k = X} \, . \qquad \text{(E)}$$

The instruments are $z_k^* = z_k \frac{H'(z_k' \beta_0)}{H(z_k' \beta_0)}$. These are "estimated" by $z_k \frac{H'\left(z_k' \hat{\beta}\right)}{H\left(z_k' \hat{\beta}\right)}$ and are equal to the z_k when *H* is the exponential function.

Properties of the method

The dissociation in a system of calibration equations (E) between the $\boldsymbol{Z}_j$ non-response explanatory variables and the $\boldsymbol{X}_j$ calibration variables results in a lower non-response bias (courtesy of the $\boldsymbol{Z}_j$) and a smaller variance (thanks to the $\boldsymbol{X}_j$).

The method requires that the number of $\boldsymbol{Z}_j$ variables (quantitative variables and indicators of quantitative variable response categories) be equal to the number of $\boldsymbol{X}_j$ (calibration) variables. In addition, the method is effective only if the correlations between the $\boldsymbol{Z}_j$ and the $\boldsymbol{X}_j$ are sufficiently strong.

Unlike the standard non-response adjustment methods, this method works even when the variables that cause the non-response are known only for respondents. In particular, it handles situations where the non-response factors are variables of interest ("non-ignorable" response mechanism).

Calmar 2 makes it possible to use this method, with the *H* functions being the usual calibration functions. Le Guennec (2004) provides an example of how the method can be applied to survey data.

REFERENCES

Caron, N. et Sautory, O. (2004). "Calages simultanés pour différentes unités d'une même enquête", *Document de travail Méthodologie statistique n° 0403, INSEE.*

Deming, W.E. and Stephan, F.F. (1940). "On a least squares adjustment of a sampled frequency table when the exact totals are known", *Annals of Mathematical Statistics*, 11, pp. 427-444.

Deville, J.-C. and Särndal, C.-E (1992). "Calibration estimation in survey sampling", .*Journal of the American Statistical Association*, 87, n°418, pp. 375-382.

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). "Generalized raking procedures in survey sampling", *Journal of the American Statistical Association*, 88, n°423, pp. 1013-1020.

Deville, J.–C. (1998). "La correction de la non-réponse par calage ou par échantillonnage équilibré", Actes du colloque de la Société Statistique du Canada, Sherbrooke, Canada.

Deville, J.–C. (2004). "La correction de la non-réponse par calage généralisé", *Actes des journées de méthodologie statistique, 16 et 17 décembre 2002, INSEE-Méthodes à paraître.*

Dupont, F. (1996). "Calage et redressement de la non-réponse totale", *Actes des journées de méthodologie statistique, 15 et 16 décembre 1993, INSEE-Méthodes n°56-57-58.*

Estevao, V. and Särndal, C.-E. (2003). "Calibration estimation in sample surveys: an overview and recent developments", article présenté au Joint Statistical Meetings de l'ASA, San Fransisco.

Fuller, W. (1987). "Measurement Error Models", New York, Wiley.

Lemaître, G. and Dufour, J. (1987). "An integrated method for weighting persons and families", *Survey Methodology*, 13, pp. 199-207.

Le Guennec, J. et Sautory, O. (2003). "La macro Calmar2, manuel d'utilisation", document interne INSEE.

Le Guennec, J. (2004). "Correction de la non-réponse par calage généralisé: une expérimentation", *Actes des journées de méthodologie statistique, 16 et 17 décembre 2002, INSEE-Méthodes à paraître.*

Lundström, S. and Särndal, C.-E. (1999). "Calibration as a standard method for treatment of nonresponse", *Journal of Official Statistics,* 15, pp. 305-327.

Roy, G. et Vanheuverzwyn, A. (2001). "Redressement par la macro CALMAR: applications et pistes d'amélioration", in *Traitements des fichiers d'enquête*. Presses Universitaires de Grenoble, pp. 31–46.

Sautory, O (1996). "Calage sur des échantillons de ménages, d'individus, d'individus-Kish, issus d'une même enquête", communication invitée aux Journées de Statistique de l'ASU, Québec, Canada.

ANNEXE

AN EXAMPLE OF SIMULTANEOUS GENERALIZED CALIBRATION

The survey

A sample has been withdrawn to investigate on population's way of life : work, level of income, cultural consumptions, social integration. The survey was performed in 1996.

About 1100 individuals have been selected through a cluster sampling design. The first stage sample includes 439 households. It is stratified according to the agglomeration size, and withdrawn by simple random sampling in the stratum, out of the previous population census, which took place in 1990. All households' members are included into the final sample.

Although the target population is composed of the individuals, that is the secondary units, the survey includes questions about the habitation and about the whole family, that is primary units, such as the number of persons in the household, the head of household's profession, the household's total income...

The calibration model

We want to calibrate the estimations on sex and age population distribution, and on households size and households professional group distribution. Households size and head of households profession are correlated both to the variables of interest and to response behaviour. Population totals in those domains are only known in the sampling frame, that is at the last census date. For that reason, the survey estimators are adjusted by calibration on those four variables totals :

- household size (single person, 2 persons…, 6 persons or more)
- head of household professional group
- individual age group
- individual sex

We also add the strata among the calibration variables in order to keep the equality between the sum of weights and the stratum population.

The update household size and head of household profession collected in the survey are supposed to be better explanatory variables for non-response than the sampling frame values. They are observed only on respondents, and their totals in the population are unknown. That's the reason why they are introduced as instrumental variables (z_k) into the calibration adjustment.

Both calibration variables and instrumental variables vectors (x_k and z_k) must have the same dimension. As we have only two real instrumental variables : update household size and head of household professional group, while the x_k vector is composed of 5 variables, we simply add to the z_k vector some of the calibration variables. This leads to the following model :

- primary units level :

$$x_{Mk} = \begin{bmatrix} \text{strata90} \\ \text{household size}\,(\text{in } 1990) \\ \text{head of household profession}\,(\text{in } 1990) \end{bmatrix} \quad z_{Mk} = \begin{bmatrix} \text{strata96}\,(= \text{strata90}) \\ \text{household size}\ \ (\text{in } 1996) \\ \text{head of household profession}\ \ (\text{in } 1996) \end{bmatrix}$$

- secondary units level :

$$x_{Ik} = \begin{bmatrix} \text{sex}\,(\text{in } 1990) \\ \text{age}\,(\text{in } 1990) \end{bmatrix} \qquad z_{Ik} = \begin{bmatrix} \text{sex_bis}\ (= \text{sex}\,) \\ \text{age_bis}\ (= \text{age}) \end{bmatrix}$$

The data sets structure

The households sample data set includes the two instrumental variables named *SEUL96* (household size in the survey) and *CS96* (head of household professional group in the survey), and the three calibration variables named *STRATA90* (strata number), *SEUL90* (household size in the sampling frame) and *CS90* (head of household professional group at the census date).

The individuals sample data set includes the two calibration variables named *AGE* and *SEX*, coming from the sampling frame.

We create the variable *STRATA96*, equal to *STRATA90*, in the households sample data set, and the variables *AGE_BIS* and *SEX_BIS* , equal to *AGE* and *SEX* , in the individuals sample data set, as pseudo-instrumental variables, in order to complete the z_k vector dimension.

The population totals data sets have the following form.

- Primary units :

var	n	r	mar1	mar2	mar3	mar4	mar5	mar6
strate90	5	0	1314	833	704	1477	777	.
seul90	2	0	3933	1172	.	.	.	.
cs90	6	0	457	470	537	435	1254	1952
strate96	5	1	.	.	.	.	.	.
seul96	2	1	.	.	.	.	.	.
cs96	6	1	.	.	.	.	.	.

- Secondary units :

var	n	r	mar1	mar2	mar3	mar4
sexe	2	0	6255	6628	.	.
age	4	0	2514	1799	5984	2586
sexe_bis	2	1	.	.	.	.
age_bis	4	1	.	.	.	.

The variable *R* points out the calibration variables (*R* =0) and the instrumental variables (*R* =1). For these ones, no population total has to be entered. In both cases, we must specify the number of levels of categorical variables (variable *N*).

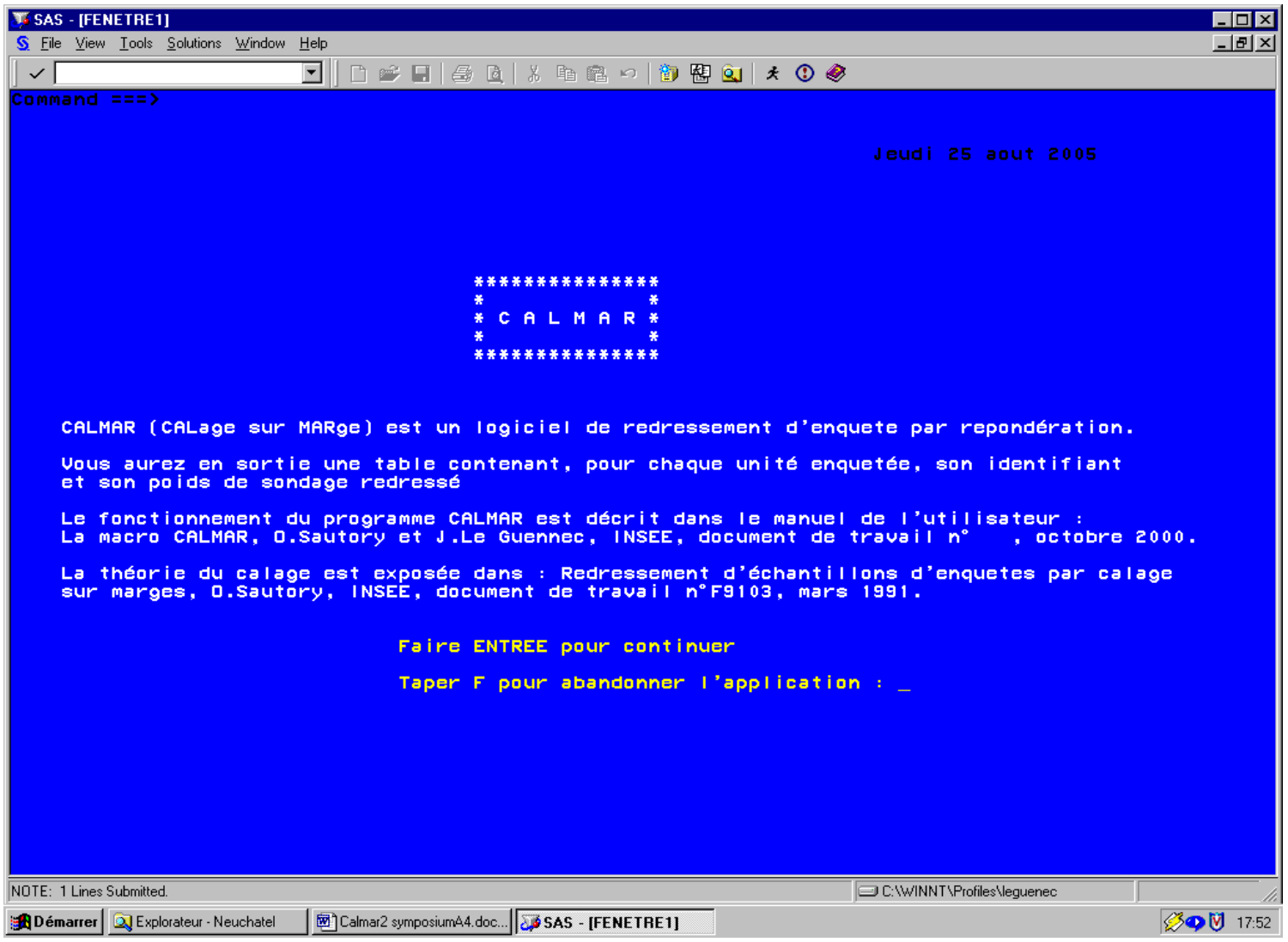
The %CALMAR2_GUIDE interface

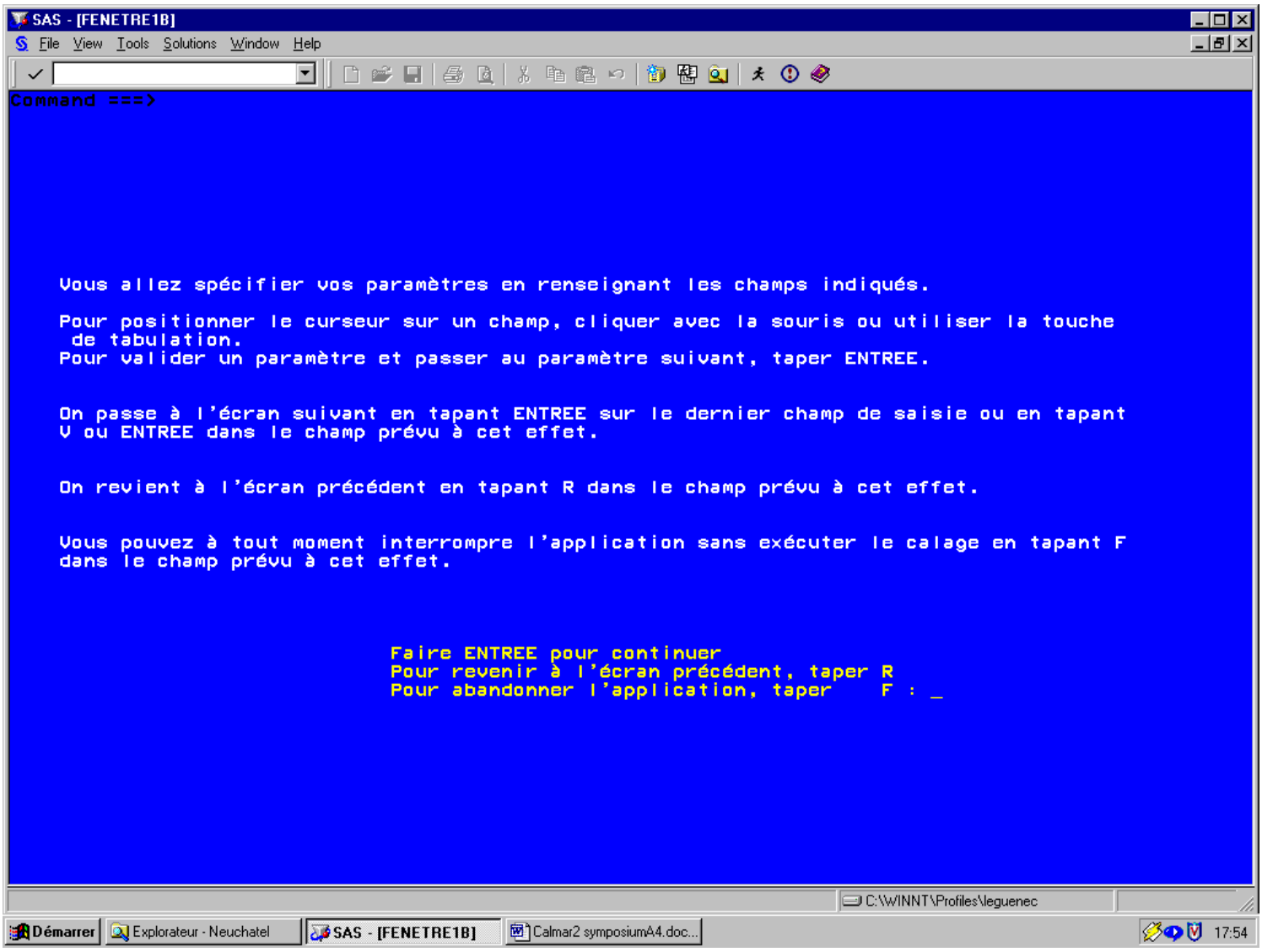
We may specify the macro parameters through the %CALMAR2_GUIDE program. The varied data files which are referred to in the calibration adjustment must have been allocated first.

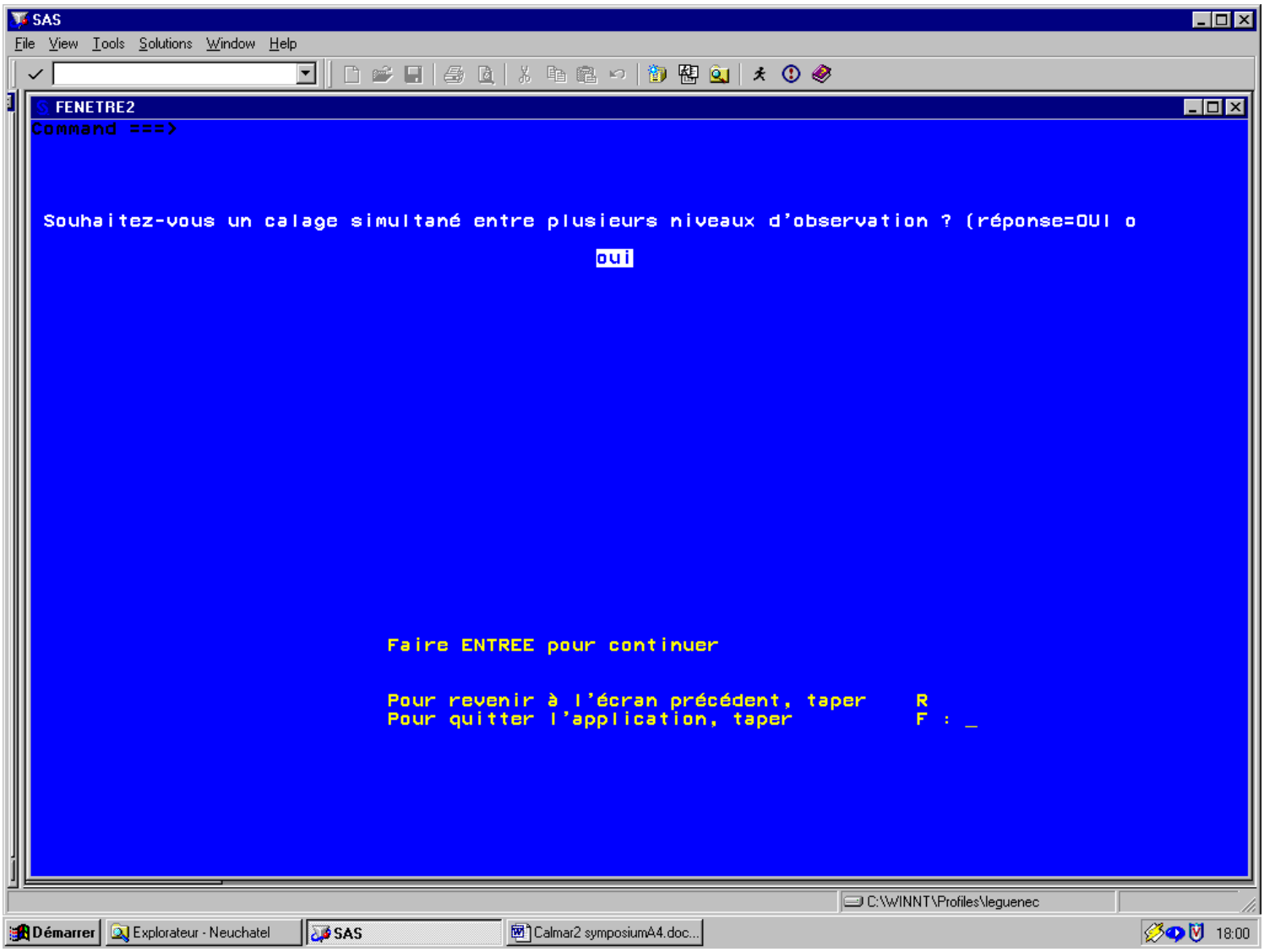
Entering the :

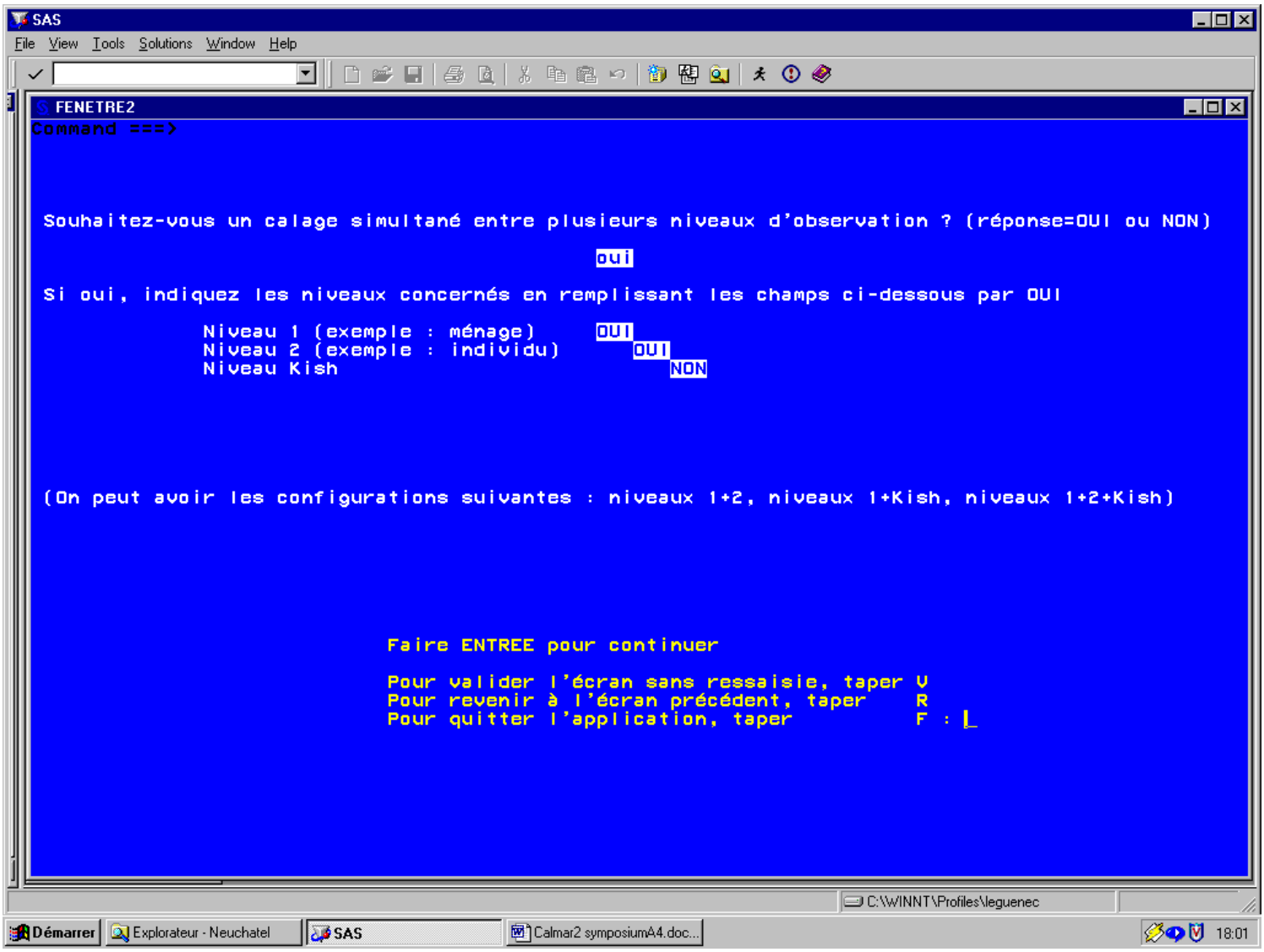
```
%CALMAR2_GUIDE
```

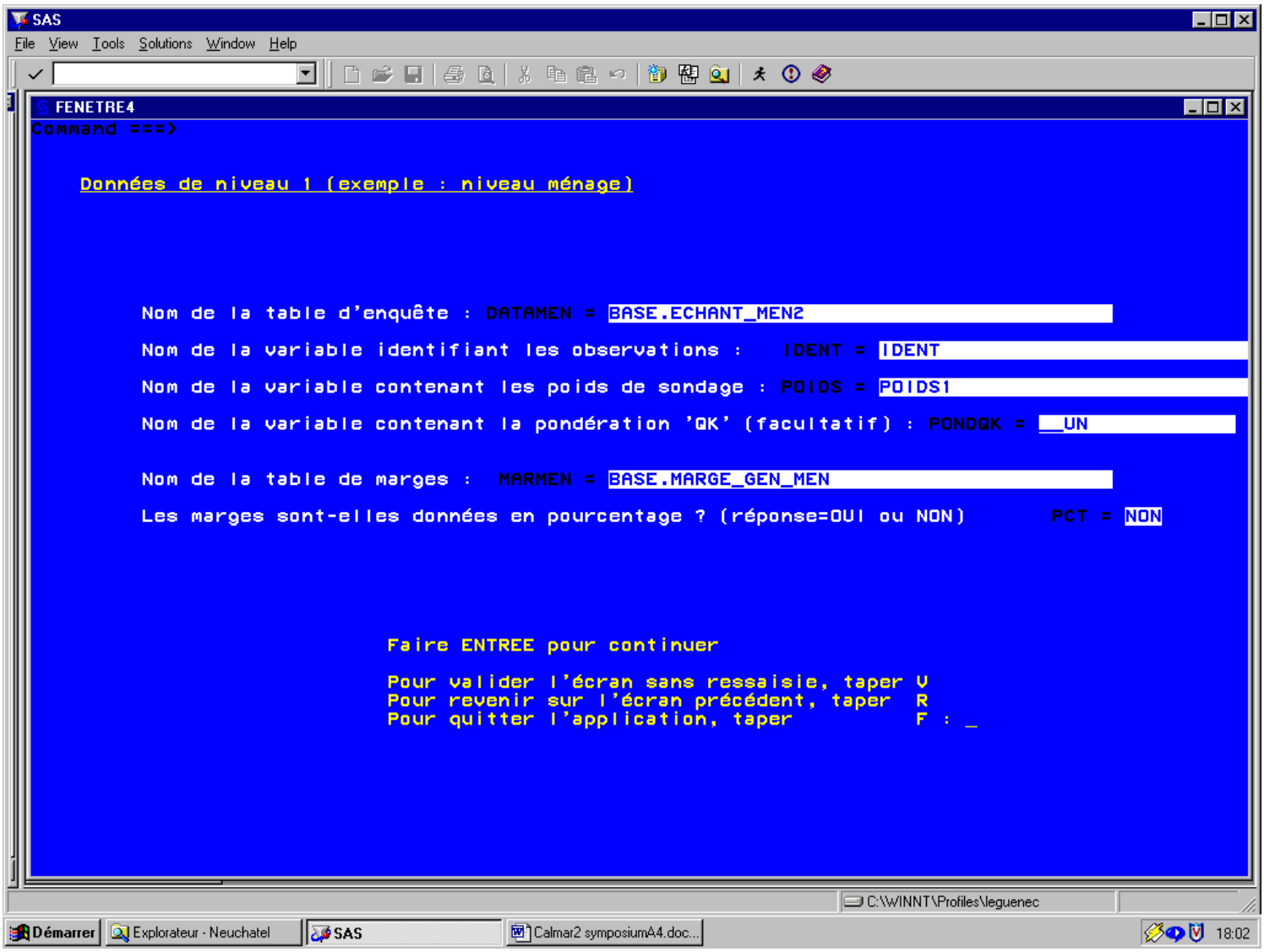
statement in the SAS Editor window makes the following interactive screens appear, which allow the user to specify his parameters value. In this example, we choose a generalized two levels simultaneous calibration.

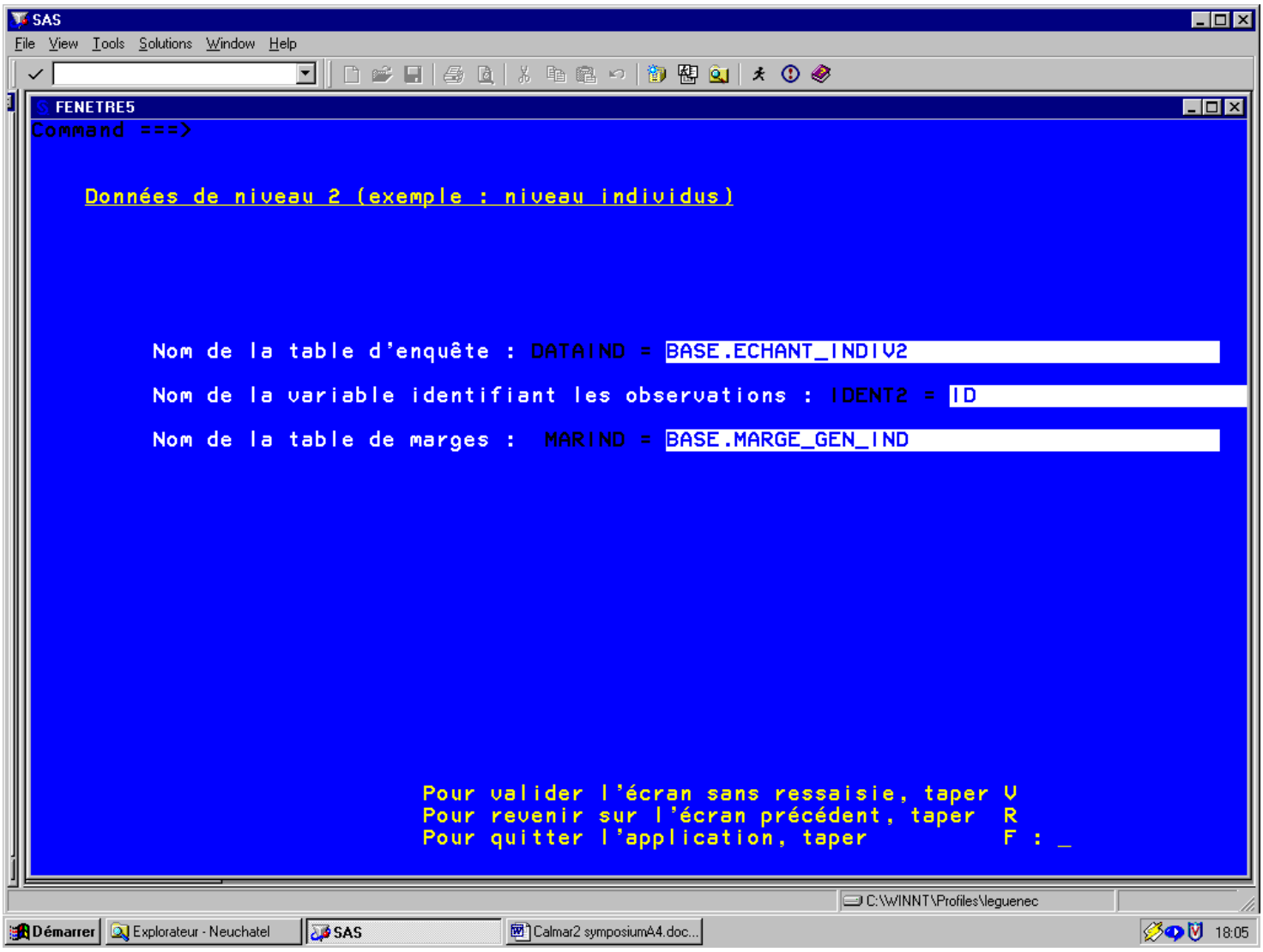

SAS - [FENETRE1]
File View Tools Solutions Window Help
Command ===>
Jeudi 25 aout 2005
* C A L M A R *
CALMAR (CALage sur MARge) est un logiciel de redressement d'enquete par repondération.
Vous aurez en sortie une table contenant, pour chaque unité enquetée, son identifiant
et son poids de sondage redressé
Le fonctionnement du programme CALMAR est décrit dans le manuel de l'utilisateur :
La macro CALMAR, O.Sautory et J.Le Guennec, INSEE, document de travail n° , octobre 2000.
La théorie du calage est exposée dans : Redressement d'échantillons d'enquetes par calage
sur marges, O.Sautory, INSEE, document de travail n°F9103, mars 1991.
Faire ENTREE pour continuer
Taper F pour abandonner l'application : _
NOTE: 1 Lines Submitted.
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
Calmar2 symposiumA4.doc...
SAS - [FENETRE1]
17:52

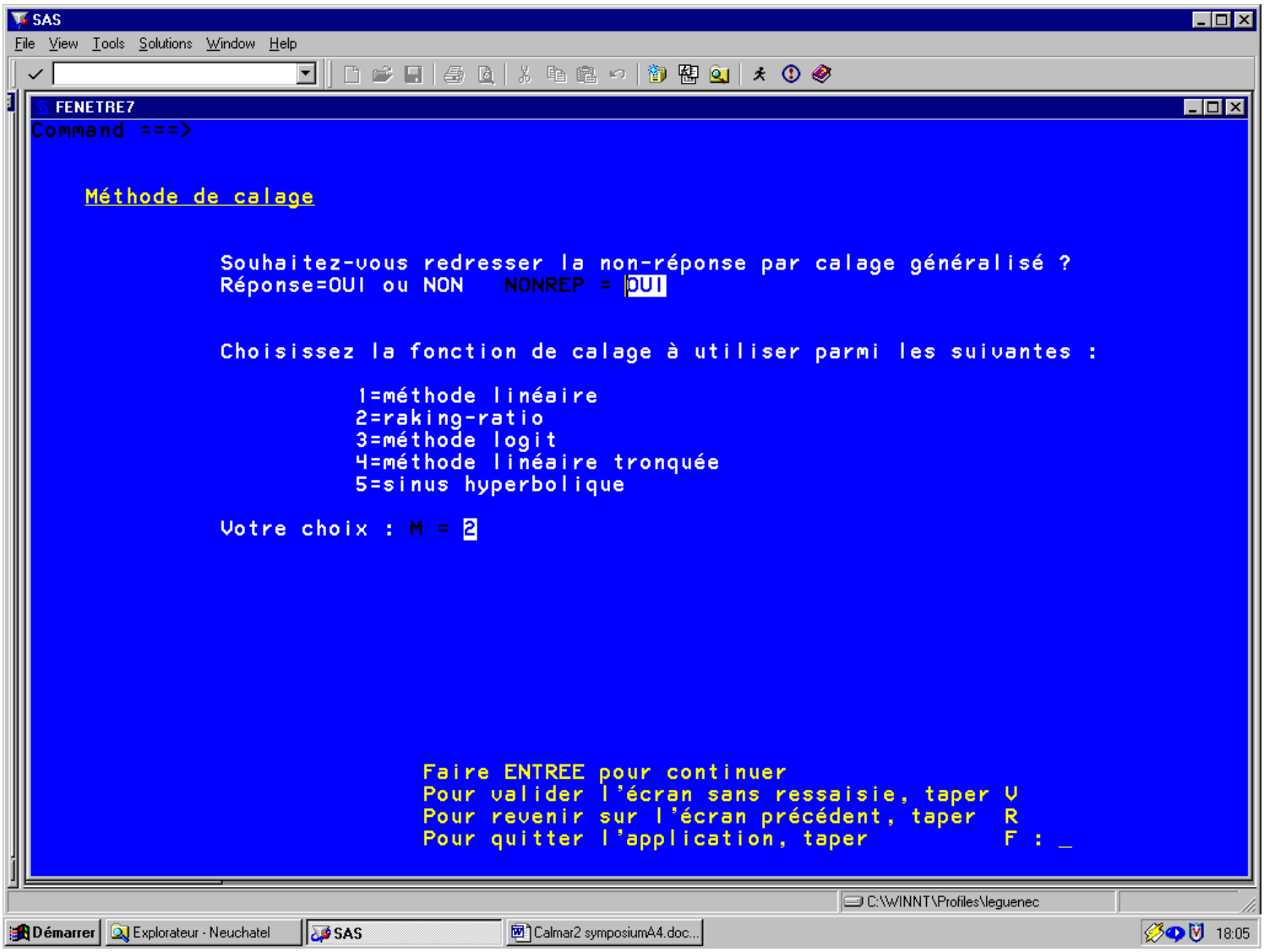

SAS - [FENETRE1B]
File View Tools Solutions Window Help
Command ===>
Vous allez spécifier vos paramètres en renseignant les champs indiqués.
Pour positionner le curseur sur un champ, cliquer avec la souris ou utiliser la touche de tabulation.
Pour valider un paramètre et passer au paramètre suivant, taper ENTREE.
On passe à l'écran suivant en tapant ENTREE sur le dernier champ de saisie ou en tapant V ou ENTREE dans le champ prévu à cet effet.
On revient à l'écran précédent en tapant R dans le champ prévu à cet effet.
Vous pouvez à tout moment interrompre l'application sans exécuter le calage en tapant F dans le champ prévu à cet effet.
Faire ENTREE pour continuer
Pour revenir à l'écran précédent, taper R
Pour abandonner l'application, taper F : _
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS - [FENETRE1B]
Calmar2 symposiumA4.doc...
17:54


SAS
File View Tools Solutions Window Help
FENETRE2
Command ===>
Souhaitez-vous un calage simultané entre plusieurs niveaux d'observation ? (réponse=OUI o
oui
Faire ENTREE pour continuer
Pour revenir à l'écran précédent, taper R
Pour quitter l'application, taper F : _
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:00

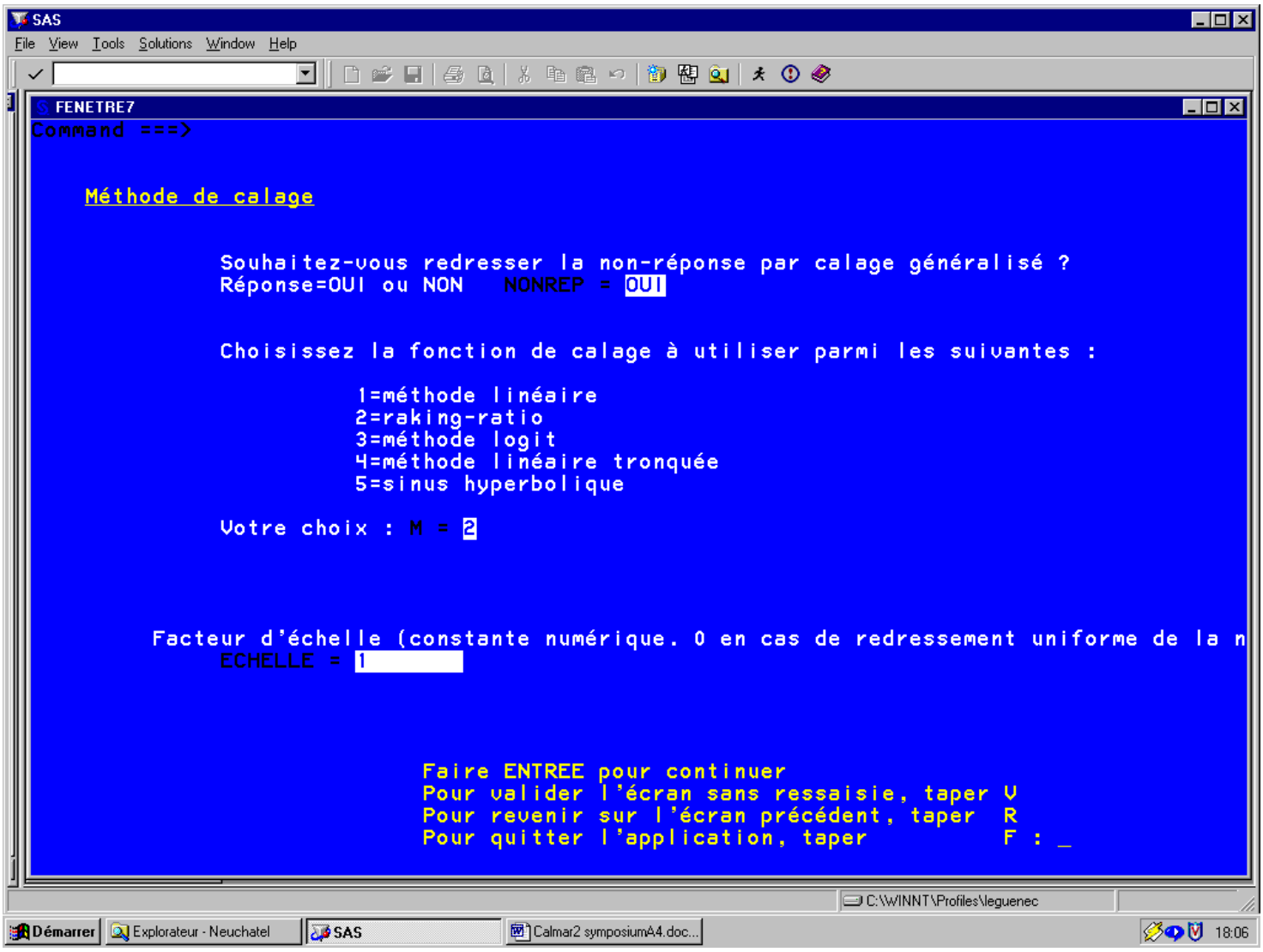

SAS
File View Tools Solutions Window Help
FENETRE2
Command ===>
Souhaitez-vous un calage simultané entre plusieurs niveaux d'observation ? (réponse=OUI ou NON)
oui
Si oui, indiquez les niveaux concernés en remplissant les champs ci-dessous par OUI
Niveau 1 (exemple : ménage) OUI
Niveau 2 (exemple : individu) OUI
Niveau Kish NON
(On peut avoir les configurations suivantes : niveaux 1+2, niveaux 1+Kish, niveaux 1+2+Kish)
Faire ENTREE pour continuer
Pour valider l'écran sans ressaisie, taper V
Pour revenir à l'écran précédent, taper R
Pour quitter l'application, taper F : L
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:01

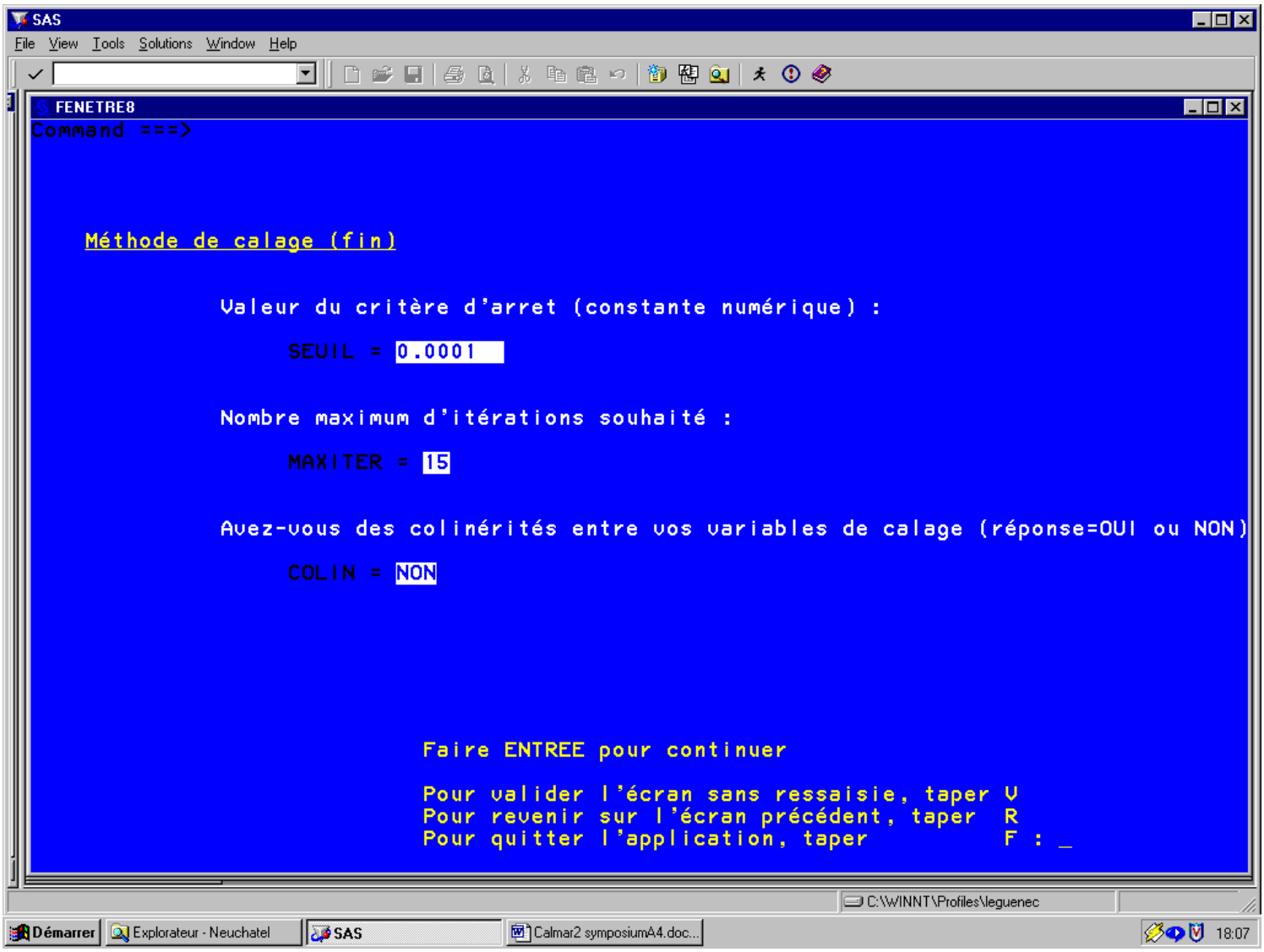

SAS
File View Tools Solutions Window Help
FENETRE4
Command ===>
Données de niveau 1 (exemple : niveau ménage)
Nom de la table d'enquête : DATAMEN = BASE.ECHANT_MEN2
Nom de la variable identifiant les observations : IDENT = IDENT
Nom de la variable contenant les poids de sondage : POIDS = POIDS1
Nom de la variable contenant la pondération 'QK' (facultatif) : PONDQK = __UN
Nom de la table de marges : MARMEN = BASE.MARGE_GEN_MEN
Les marges sont-elles données en pourcentage ? (réponse=OUI ou NON) PCT = NON
Faire ENTREE pour continuer
Pour valider l'écran sans ressaisie, taper V
Pour revenir sur l'écran précédent, taper R
Pour quitter l'application, taper F : _
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:02


SAS
File View Tools Solutions Window Help
FENETRE5
Command ===>
Données de niveau 2 (exemple : niveau individus)
Nom de la table d'enquête : DATAIND = BASE.ECHANT_INDIV2
Nom de la variable identifiant les observations : IDENT2 = ID
Nom de la table de marges : MARIND = BASE.MARGE_GEN_IND
Pour valider l'écran sans ressaisie, taper V
Pour revenir sur l'écran précédent, taper R
Pour quitter l'application, taper F : _
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:05

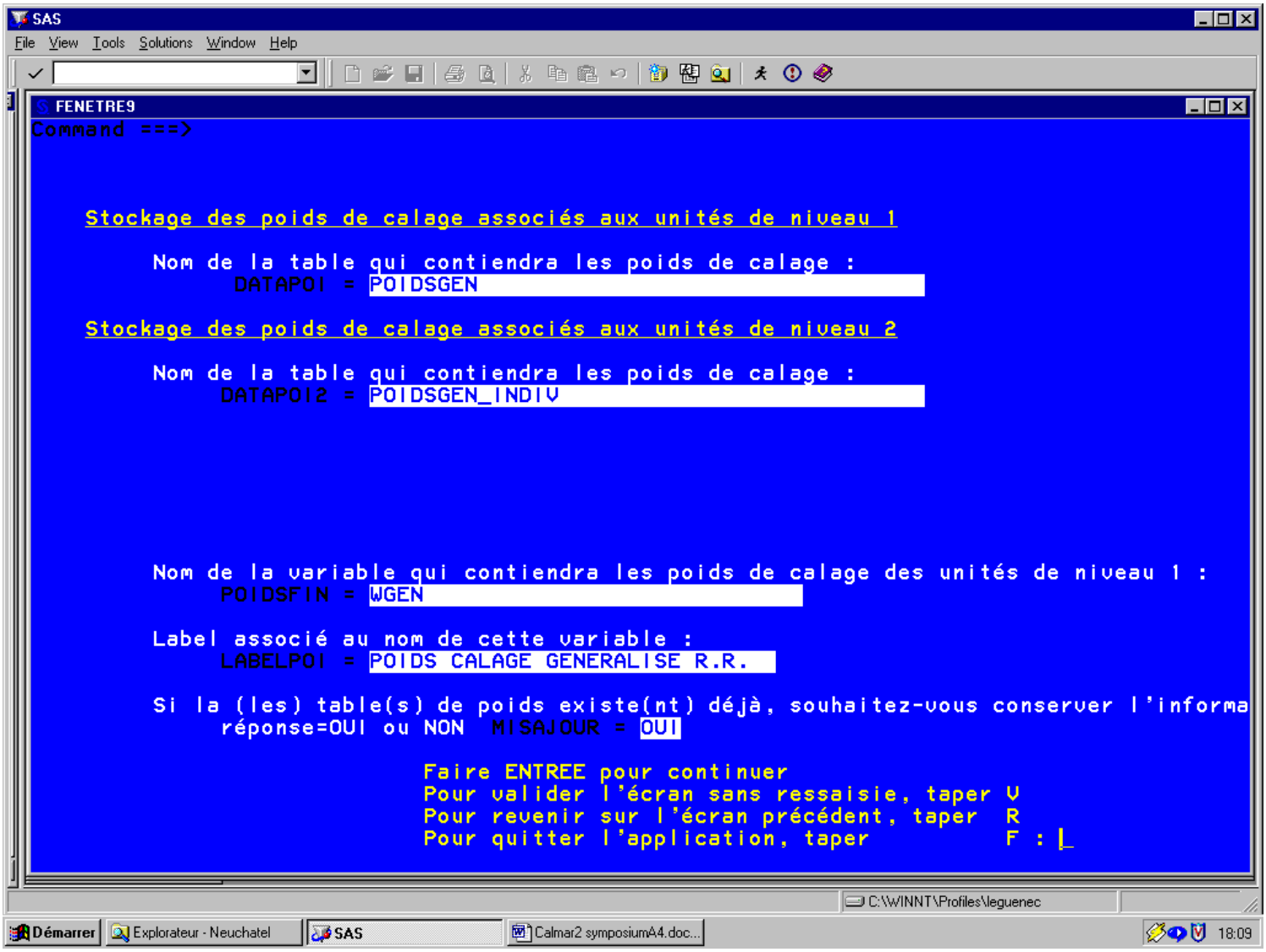


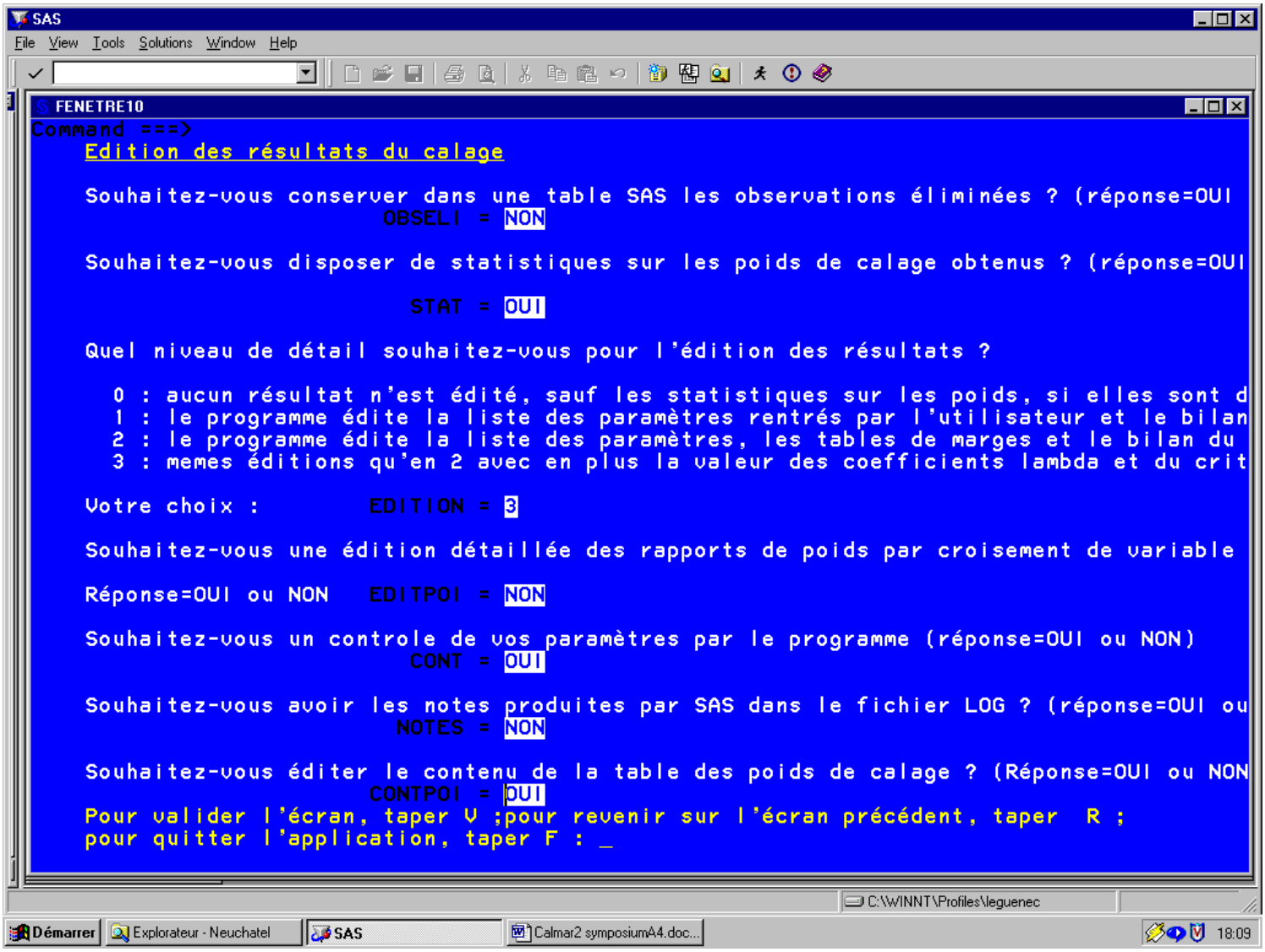
Using the parameter : NONREP=OUI is the way to perform a generalized calibration.

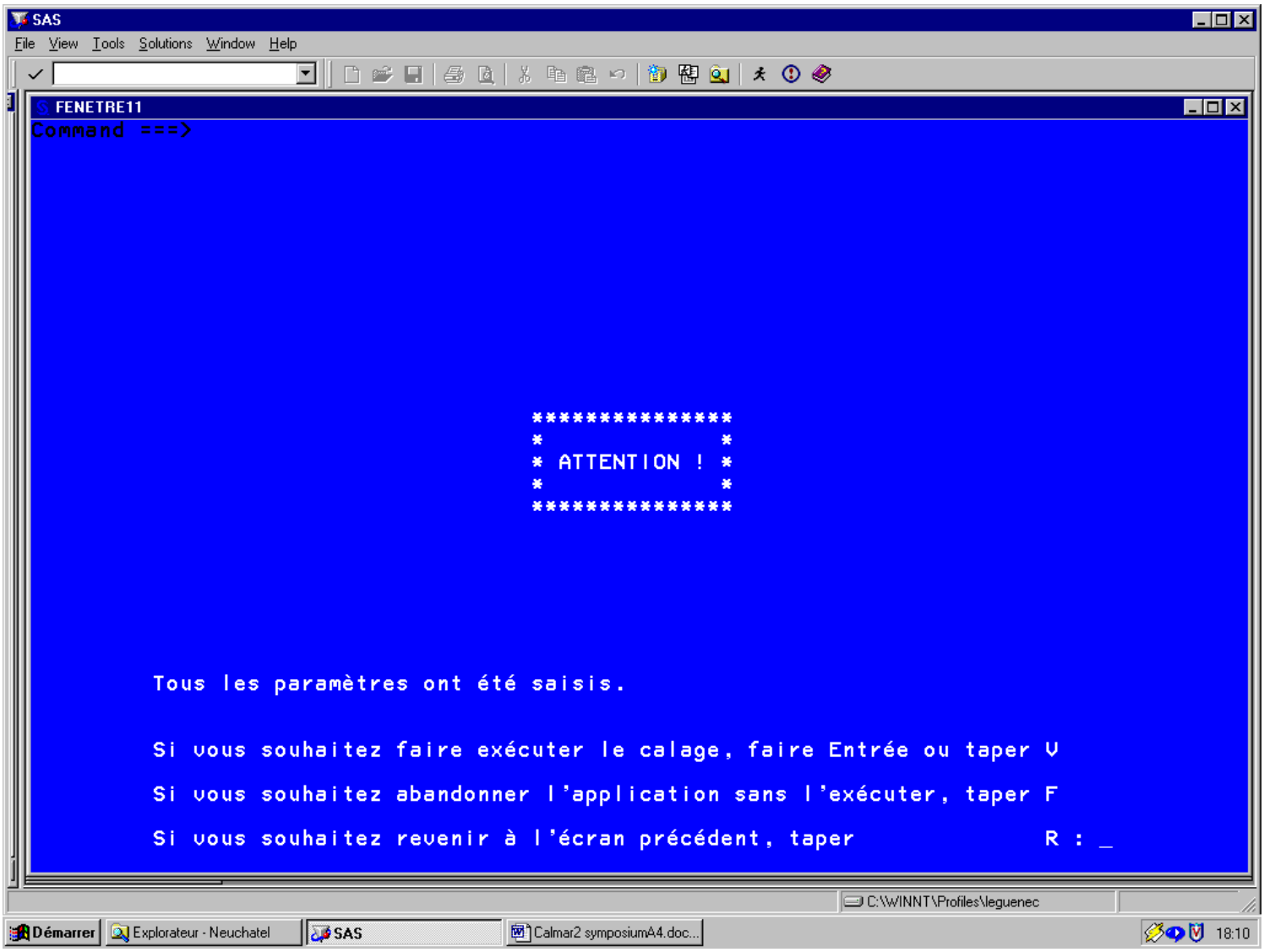

SAS
File View Tools Solutions Window Help
FENETRE7
Command ===>
Méthode de calage
Souhaitez-vous redresser la non-réponse par calage généralisé ?
Réponse=OUI ou NON NONREP = OUI
Choisissez la fonction de calage à utiliser parmi les suivantes :
1=méthode linéaire
2=raking-ratio
3=méthode logit
4=méthode linéaire tronquée
5=sinus hyperbolique
Votre choix : M = 2
Facteur d'échelle (constante numérique. 0 en cas de redressement uniforme de la n
ECHELLE = 1
Faire ENTREE pour continuer
Pour valider l'écran sans ressaisie, taper V
Pour revenir sur l'écran précédent, taper R
Pour quitter l'application, taper F : _
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:06



In case of failure of a first calibration program because of hidden colinearities among calibration variables, you may change the COLIN parameter value into OUI, which involves the use of generalized inverse matrices.


SAS
File View Tools Solutions Window Help
FENETRE9
Command ===>
Stockage des poids de calage associés aux unités de niveau 1
Nom de la table qui contiendra les poids de calage :
DATAPOI = POIDSGEN
Stockage des poids de calage associés aux unités de niveau 2
Nom de la table qui contiendra les poids de calage :
DATAPOI2 = POIDSGEN_INDIV
Nom de la variable qui contiendra les poids de calage des unités de niveau 1 :
POIDSFIN = WGEN
Label associé au nom de cette variable :
LABELPOI = POIDS CALAGE GENERALISE R.R.
Si la (les) table(s) de poids existe(nt) déjà, souhaitez-vous conserver l'informa
réponse=OUI ou NON MISAJOUR = OUI
Faire ENTREE pour continuer
Pour valider l'écran sans ressaisie, taper V
Pour revenir sur l'écran précédent, taper R
Pour quitter l'application, taper F :
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:09


SAS
File View Tools Solutions Window Help
FENETRE10
Command ===>
Edition des résultats du calage
Souhaitez-vous conserver dans une table SAS les observations éliminées ? (réponse=OUI
OBSELI = NON
Souhaitez-vous disposer de statistiques sur les poids de calage obtenus ? (réponse=OUI
STAT = OUI
Quel niveau de détail souhaitez-vous pour l'édition des résultats ?
0 : aucun résultat n'est édité, sauf les statistiques sur les poids, si elles sont d
1 : le programme édite la liste des paramètres rentrés par l'utilisateur et le bilan
2 : le programme édite la liste des paramètres, les tables de marges et le bilan du
3 : memes éditions qu'en 2 avec en plus la valeur des coefficients lambda et du crit
Votre choix : EDITION = 3
Souhaitez-vous une édition détaillée des rapports de poids par croisement de variable
Réponse=OUI ou NON EDITPOI = NON
Souhaitez-vous un controle de vos paramètres par le programme (réponse=OUI ou NON)
CONT = OUI
Souhaitez-vous avoir les notes produites par SAS dans le fichier LOG ? (réponse=OUI ou
NOTES = NON
Souhaitez-vous éditer le contenu de la table des poids de calage ? (Réponse=OUI ou NON
CONTPOI = OUI
Pour valider l'écran, taper V ;pour revenir sur l'écran précédent, taper R ;
pour quitter l'application, taper F : _
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:09


SAS
File View Tools Solutions Window Help
FENETRE11
Command ===>
****
* *
* ATTENTION ! *
* *
****
Tous les paramètres ont été saisis.
Si vous souhaitez faire exécuter le calage, faire Entrée ou taper V
Si vous souhaitez abandonner l'application sans l'exécuter, taper F
Si vous souhaitez revenir à l'écran précédent, taper R : _
C:\WINNT\Profiles\leguenec
Démarrer
Explorateur - Neuchatel
SAS
Calmar2 symposiumA4.doc...
18:10

This program has produced the following output :

```
                    *********************************
                    ***   PARAMÈTRES DE LA MACRO   ***
                    *********************************


TABLE(S) EN ENTRÉE :
 TABLE DE DONNÉES DE NIVEAU 1             DATAMEN   =  BASE.ECHANT_MEN2
    IDENTIFIANT DU NIVEAU 1               IDENT     =  IDENT
 TABLE DE DONNÉES DE NIVEAU 2             DATAIND   =  BASE.ECHANT_INDIV2
    IDENTIFIANT DU NIVEAU 2               IDENT2    =  ID
 TABLE DES INDIVIDUS KISH                 DATAKISH  =
 PONDÉRATION INITIALE                     POIDS     =  POIDS1
 FACTEUR D'ÉCHELLE                        ECHELLE   =  1
 PONDÉRATION QK                           PONDQK    =  __UN
 PONDÉRATION KISH                         POIDKISH  =

TABLE(S) DES MARGES :
 DE NIVEAU 1                              MARMEN    =  BASE.MARGE_GEN_MEN
 DE NIVEAU 2                              MARIND    =  BASE.MARGE_GEN_IND
 DE NIVEAU KISH                           MARKISH   =
 MARGES EN POURCENTAGES                   PCT       =  NON
 EFFECTIF DANS LA POPULATION :
  DES ÉLÉMENTS DE NIVEAU 1                POPMEN    =
  DES ÉLÉMENTS DE NIVEAU 2                POPIND    =
  DES ÉLÉMENTS KISH                       POPKISH   =

REDRESSEMENT DE LA NON-RÉPONSE DEMANDÉ :  NONREP    =  OUI

MÉTHODE UTILISÉE                          M         =  2
 BORNE INFÉRIEURE                         LO        =
 BORNE SUPÉRIEURE                         UP        =
 COEFFICIENT DU SINUS HYPERBOLIQUE        ALPHA     =
 SEUIL D'ARRÊT                            SEUIL     =  0.0001
 NOMBRE MAXIMUM D'ITÉRATIONS              MAXITER   =  15
 TRAITEMENT DES COLINÉARITÉS              COLIN     =  NON

TABLE(S) CONTENANT LA POND. FINALE
 DE NIVEAU 1                              DATAPOI       =  POIDSGEN
 DE NIVEAU 2                              DATAPOI2      =  POIDSGEN_INDIV
 DE NIVEAU KISH                           DATAPOI3      =
 MISE À JOUR DE(S) TABLE(S) DATAPOI(2)(3) MISAJOUR      =  OUI
 PONDÉRATION FINALE DES UNITES 1 ET 2     POIDSFIN      =  WGEN
 LABEL DE LA PONDÉRATION FINALE           LABELPOI      =  POIDS CALAGE GENERALISE R.R.
 PONDÉRATION FINALE DES UNITES KISH       POIDSKISHFIN  =
 LABEL DE LA PONDÉRATION KISH             LABELPOIKISH  =
 CONTENU DE(S) TABLE(S) DATAPOI(2)(3)     CONTPOI       =  OUI

ÉDITION DES RÉSULTATS                     EDITION       =  3
 ÉDITION DES POIDS                        EDITPOI       =  NON
 STATISTIQUES SUR LES POIDS               STAT          =  OUI

CONTRÔLES                                 CONT          =  OUI

TABLE CONTENANT LES OBS. ÉLIMINÉES        OBSELI        =  NON
NOTES SAS                                 NOTES         =  NON
```

COMPARAISON ENTRE LES MARGES TIRÉES DE L'ÉCHANTILLON (AVEC LA PONDÉRATION INITIALE)
ET LES MARGES DANS LA POPULATION (MARGES DU CALAGE)

VARIABLE	MODALITÉ	MARGE ÉCHANTILLON	MARGE POPULATION	POURCENTAGE ÉCHANTILLON	POURCENTAGE POPULATION
CS90	2	474.87	457	9.30	8.95
	3	548.05	470	10.73	9.21
	4	452.79	537	8.87	10.52
	5	408.56	435	8.00	8.52
	6	1335.78	1254	26.16	24.56
	7	1886.68	1952	36.94	38.24
SEUL90	0	3994.25	3933	78.22	77.04
	1	1112.48	1172	21.78	22.96
STRATE90	0	1336.76	1314	26.18	25.74
	1	850.08	833	16.65	16.32
	2	668.15	704	13.08	13.79
	3	1452.54	1477	28.44	28.93
	4	799.20	777	15.65	15.22
AGE	00-14 an	2714.64	2514	20.44	19.51
	15-24 an	1859.50	1799	14.00	13.96
	25-59 an	5853.04	5984	44.06	46.45
	60- ? an	2856.77	2586	21.51	20.07
SEXE	1	6610.65	6255	49.76	48.55
	2	6673.30	6628	50.24	51.45

MÉTHODE : RAKING RATIO
PREMIER TABLEAU RÉCAPITULATIF DE L'ALGORITHME :
LA VALEUR DU CRITÈRE D'ARRÊT ET LE NOMBRE DE POIDS NÉGATIFS APRÈS CHAQUE ITÉRATION

ITÉRATION	CRITÈRE D'ARRÊT	POIDS NÉGATIFS
1	31.8702	0
2	19.4266	0
3	6.6239	0
4	1.6661	0
5	0.1647	0
6	0.0018	0
7	0.0000	0

MÉTHODE : RAKING RATIO
DEUXIÈME TABLEAU RÉCAPITULATIF DE L'ALGORITHME :
LES COEFFICIENTS DU VECTEUR LAMBDA DE MULTIPLICATEURS DE LAGRANGE APRÈS CHAQUE ITÉRATION

VARIABLE	MODALITÉ	LAMBDA1	LAMBDA2	LAMBDA3	LAMBDA4	LAMBDA5	LAMBDA6	LAMBDA7
CS90	2	-1.72987	-1.06367	-0.74162	-0.59605	-0.58433	-0.58423	-0.58423
CS90	3	-0.73203	-0.81548	-0.71064	-0.66919	-0.66734	-0.66734	-0.66734
CS90	4	-1.52491	-0.90137	-0.57274	-0.46619	-0.45885	-0.45880	-0.45880
CS90	5	0.02398	-0.08362	0.05495	0.15849	0.17125	0.17139	0.17139
CS90	6	-1.44313	-1.27743	-1.16253	-1.16950	-1.17952	-1.17978	-1.17978
CS90	7	2.01767	1.44720	1.08231	0.93289	0.91463	0.91443	0.91443
SEUL90	0	0.28744	0.24031	0.10618	0.01675	0.00398	0.00382	0.00382
SEUL90	1	.	.	.	.	.	.	.
STRATE90	0	-0.00359	-0.03815	-0.05350	-0.06461	-0.06705	-0.06708	-0.06708
STRATE90	1	-0.07570	-0.05431	-0.01510	-0.00174	-0.00141	-0.00142	-0.00142
STRATE90	2	0.02896	-0.02652	-0.04203	-0.05293	-0.05605	-0.05611	-0.05611
STRATE90	3	0.15277	0.09427	0.05985	0.04150	0.03791	0.03785	0.03785
STRATE90	4	.	.	.	.	.	.	.
AGE	00-14 an	-0.04781	-0.05975	-0.05073	-0.04011	-0.03860	-0.03858	-0.03858
AGE	15-24 an	-0.13633	-0.14540	-0.10953	-0.07838	-0.07366	-0.07360	-0.07360
AGE	25-59 an	0.58023	0.45016	0.37013	0.35722	0.35799	0.35802	0.35802
AGE	60- ? an	-1.24146	-1.00060	-0.73525	-0.60728	-0.59095	-0.59075	-0.59075
SEXE	1	-0.12577	-0.08359	-0.06877	-0.06586	-0.06527	-0.06526	-0.06526
SEXE	2	.	.	.	.	.	.	.

MÉTHODE : RAKING RATIO
COMPARAISON ENTRE LES MARGES FINALES DANS L'ÉCHANTILLON (AVEC LA PONDÉRATION FINALE)
ET LES MARGES DANS LA POPULATION (MARGES DU CALAGE)

VARIABLE	MODALITÉ	MARGE ÉCHANTILLON	MARGE POPULATION	POURCENTAGE ÉCHANTILLON	POURCENTAGE POPULATION
CS90	2	457	457	8.95	8.95
	3	470	470	9.21	9.21
	4	537	537	10.52	10.52
	5	435	435	8.52	8.52
	6	1254	1254	24.56	24.56
	7	1952	1952	38.24	38.24
SEUL90	0	3933	3933	77.04	77.04
	1	1172	1172	22.96	22.96
STRATE90	0	1314	1314	25.74	25.74
	1	833	833	16.32	16.32
	2	704	704	13.79	13.79
	3	1477	1477	28.93	28.93
	4	777	777	15.22	15.22
AGE	00-14 an	2514	2514	19.51	19.51
	15-24 an	1799	1799	13.96	13.96
	25-59 an	5984	5984	46.45	46.45
	60- ? an	2586	2586	20.07	20.07
SEXE	1	6255	6255	48.55	48.55
	2	6628	6628	51.45	51.45

MÉTHODE : RAKING RATIO
STATISTIQUES SUR LES RAPPORTS DE POIDS (= PONDÉRATIONS FINALES / PONDÉRATIONS INITIALES)
ET SUR LES PONDÉRATIONS FINALES

The UNIVARIATE Procedure
Variable: _F_ (RAPPORT DE POIDS)

Basic Statistical Measures

Location		Variability	
Mean	1.000153	Std Deviation	0.63381
Median	0.841701	Variance	0.40172
Mode	0.718970	Range	4.84255
		Interquartile Range	0.64433

Quantiles (Definition 5)

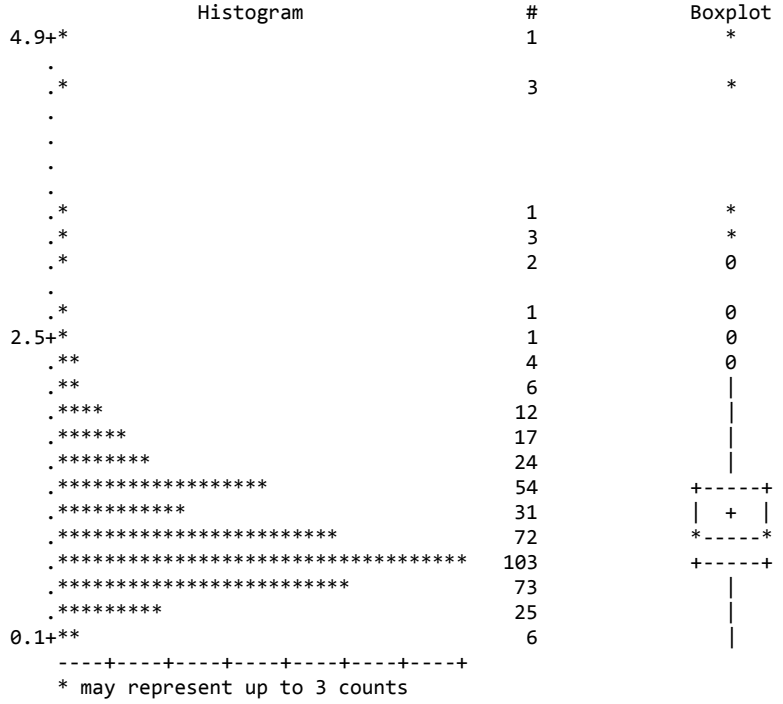
Quantile	Estimate
100% Max	4.987209
99%	3.442740
95%	2.055925
90%	1.680763
75% Q3	1.292517
50% Median	0.841701
25% Q1	0.648192
10%	0.427769
5%	0.373577
1%	0.176809
0% Min	0.144655

Extreme Observations

----------Lowest----------			----------Highest----------		
Value	IDENT	Obs	Value	IDENT	Obs
0.144655	9363006020	413	3.44274	2163019030	85
0.160953	7269012040	311	4.49041	2369009020	118
0.160953	5369013020	280	4.49041	8269014980	366
0.160953	2169020050	101	4.53997	5363003760	254
0.176809	9363022260	420	4.98721	9363033000	425

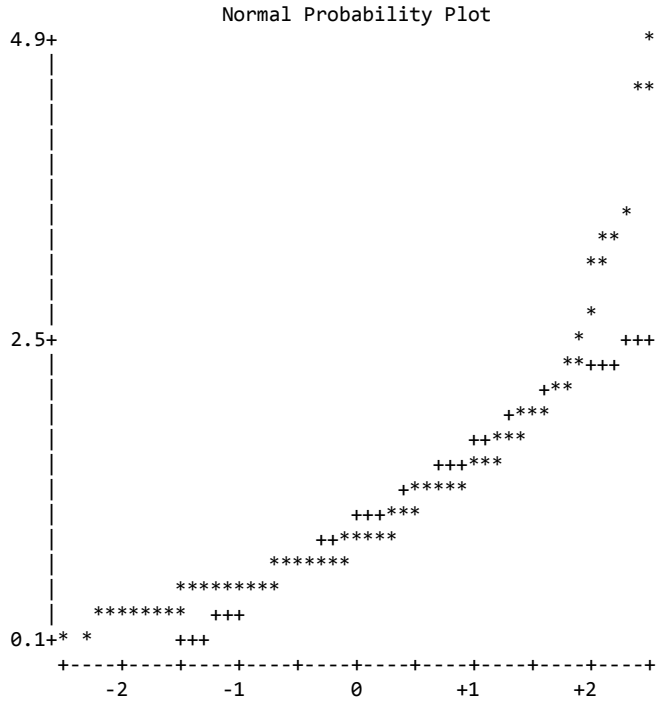
MÉTHODE : RAKING RATIO
STATISTIQUES SUR LES RAPPORTS DE POIDS (= PONDÉRATIONS FINALES / PONDÉRATIONS INITIALES)
ET SUR LES PONDÉRATIONS FINALES

The UNIVARIATE Procedure
Variable: _F_ (RAPPORT DE POIDS)



MÉTHODE : RAKING RATIO
STATISTIQUES SUR LES RAPPORTS DE POIDS (= PONDÉRATIONS FINALES / PONDÉRATIONS INITIALES)
ET SUR LES PONDÉRATIONS FINALES

The UNIVARIATE Procedure
Variable: _F_ (RAPPORT DE POIDS)



MÉTHODE : RAKING RATIO
STATISTIQUES SUR LES RAPPORTS DE POIDS (= PONDÉRATIONS FINALES / PONDÉRATIONS INITIALES)
ET SUR LES PONDÉRATIONS FINALES

The UNIVARIATE Procedure
Variable: __WFIN (PONDÉRATION FINALE)

Basic Statistical Measures

Location		Variability	
Mean	11.62870	Std Deviation	8.61949
Median	9.78269	Variance	74.29559
Mode	6.94525	Range	82.93829
		Interquartile Range	7.63255

Quantiles (Definition 5)

Quantile	Estimate
100% Max	84.23396
99%	44.74633
95%	25.32619
90%	21.86779
75% Q3	14.02605
50% Median	9.78269
25% Q1	6.39351
10%	4.43868
5%	3.62805
1%	2.08119
0% Min	1.29567

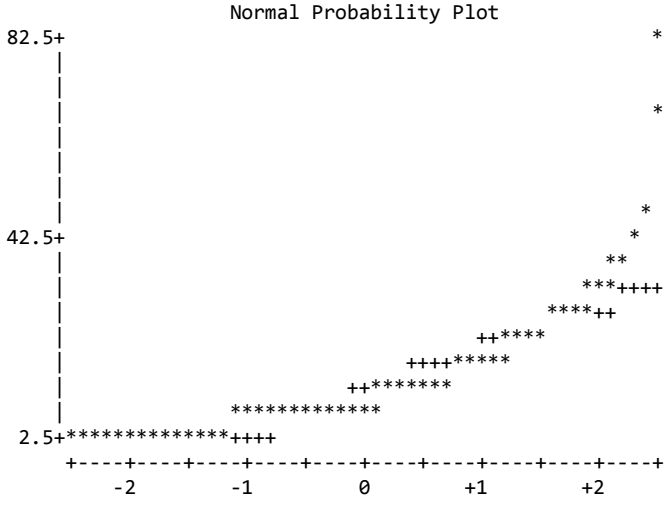
Extreme Observations

----------Lowest----------			----------Highest----------		
Value	IDENT	Obs	Value	IDENT	Obs
1.29567	7269012040	311	44.7463	3169012010	187
1.29567	5369013020	280	48.5008	2163000830	77
1.29567	2169020050	101	65.2456	2369009020	118
2.03368	1163023120	16	65.2456	8269014980	366
2.08119	8369002830	386	84.2340	9363033000	425

MÉTHODE : RAKING RATIO
STATISTIQUES SUR LES RAPPORTS DE POIDS (= PONDÉRATIONS FINALES / PONDÉRATIONS INITIALES)
ET SUR LES PONDÉRATIONS FINALES

The UNIVARIATE Procedure
Variable: __WFIN (PONDÉRATION FINALE)

```
                    Histogram                          #             Boxplot
 82.5+*                                                1                *
     .
     .
     .*                                                2                *
     .
     .
     .
     .*                                                1                *
 42.5+*                                                1                *
     .*                                                3                0
     .*                                                4                0
     .****                                            14                0
     .******                                          23                |
     .*************                                   50                |
     .************************                        99               +--+--+
     .*********************************************  185               *-----*
  2.5+**************                                   56                |
     ----+----+----+----+----+----+----+----+----+--
     * may represent up to 4 counts
```



MÉTHODE : RAKING RATIO
RAPPORTS DE POIDS MOYENS (PONDÉRATIONS FINALES / PONDÉRATIONS INITIALES)
POUR CHAQUE VALEUR DES VARIABLES

VARIABLE	MODALITE	NOMBRE D'OBSERVATIONS DE NIVEAU 1	RAPPORT DE POIDS
CS90	2	38	0.94738
CS90	3	51	0.88555
CS90	4	39	1.12938
CS90	5	39	1.07181
CS90	6	113	0.91558
CS90	7	159	1.06035
SEUL90	0	341	0.98348
SEUL90	1	98	1.05817
STRATE90	0	92	0.98297
STRATE90	1	88	0.97991
STRATE90	2	83	1.05366
STRATE90	3	86	1.01684
STRATE90	4	90	0.97222
ENSEMBLE		439	1.00015

MÉTHODE : RAKING RATIO
RAPPORTS DE POIDS MOYENS (PONDÉRATIONS FINALES / PONDÉRATIONS INITIALES)
POUR CHAQUE VALEUR DES VARIABLES

VARIABLE	MODALITE	NOMBRE D'OBSERVATIONS DE NIVEAU 2	RAPPORT DE POIDS
AGE	00-14 an	230	0.91079
AGE	15-24 an	155	0.96502
AGE	25-59 an	506	1.02352
AGE	60- ? an	239	0.91740
SEXE	1	561	0.94519
SEXE	2	569	0.99468
ENSEMBLE		1130	0.97011

MÉTHODE : RAKING RATIO
CONTENU DE LA TABLE POIDSGEN CONTENANT LA NOUVELLE PONDÉRATION WGEN

The CONTENTS Procedure

Data Set Name:	WORK.POIDSGEN	Observations:	439
Member Type:	DATA	Variables:	2
Engine:	V8	Indexes:	0
Created:	18:10 Thursday, August 25, 2005	Observation Length:	24
Last Modified:	18:10 Thursday, August 25, 2005	Deleted Observations:	0
Protection:		Compressed:	NO
Data Set Type:		Sorted:	NO
Label:			

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos	Label
1	IDENT	Char	10	8	
2	WGEN	Num	8	0	POIDS CALAGE GENERALISE R.R.

MÉTHODE : RAKING RATIO
CONTENU DE LA TABLE POIDSGEN_INDIV CONTENANT LA NOUVELLE PONDÉRATION WGEN

The CONTENTS Procedure

Data Set Name:	WORK.POIDSGEN_INDIV	Observations:	1130
Member Type:	DATA	Variables:	3
Engine:	V8	Indexes:	0
Created:	18:10 Thursday, August 25, 2005	Observation Length:	32
Last Modified:	18:10 Thursday, August 25, 2005	Deleted Observations:	0
Protection:		Compressed:	NO
Data Set Type:		Sorted:	NO
Label:			

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos	Label
1	IDENT	Char	10	8	
3	WGEN	Num	8	0	POIDS CALAGE GENERALISE R.R.
2	id	Char	12	18	

```
                ********************
                ***     BILAN     ***
                ********************
*
*   DATE : 25 AOUT 2005              HEURE : 17:48
*
*   ***********************************
*   TABLE EN ENTRÉE : BASE.ECHANT_MEN2
*   ***********************************
*
*   NOMBRE D'OBSERVATIONS DANS LA TABLE EN ENTRÉE  :            439
*   NOMBRE D'OBSERVATIONS ÉLIMINÉES                :            0
*   NOMBRE D'OBSERVATIONS CONSERVÉES               :            439
*
*   VARIABLE DE PONDÉRATION : POIDS1
*
*   NOMBRE DE VARIABLES CATÉGORIELLES : 3
*   LISTE DES VARIABLES CATÉGORIELLES ET DE LEURS NOMBRES DE MODALITÉS :
      cs90 (6) seul90 (2) strate90 (5)
*
*   SOMME DES POIDS INITIAUX                                  : 5107
*   TAILLE DE LA POPULATION                                   : 5105
*
* VARIABLES DE NON-REPONSE
*    NOMBRE DE VARIABLES CATEGORIELLES : 3
*    LISTE DES VARIABLES CATÉGORIELLES ET DE LEURS NOMBRES DE MODALITÉS :
      cs96 (6)  seul96 (2)  strate96 (5)
*
*   ***********************************
*   TABLE EN ENTRÉE : BASE.ECHANT_INDIV2
*   ***********************************
*
*   NOMBRE D'OBSERVATIONS DANS LA TABLE EN ENTRÉE  :            1130
*   NOMBRE D'OBSERVATIONS ÉLIMINÉES                :            0
*   NOMBRE D'OBSERVATIONS CONSERVÉES               :            1130
*
*   NOMBRE DE VARIABLES CATÉGORIELLES : 2
*   LISTE DES VARIABLES CATÉGORIELLES ET DE LEURS NOMBRES DE MODALITÉS :
*     age (4) sexe (2)

*   SOMME DES POIDS INITIAUX                                  : 13284
*   TAILLE DE LA POPULATION                                   : 12883
*
* VARIABLES DE NON-REPONSE
*    NOMBRE DE VARIABLES CATEGORIELLES : 2
*    LISTE DES VARIABLES CATÉGORIELLES ET DE LEURS NOMBRES DE MODALITÉS :
      age_bis (4)  sexe_bis (2)
*
*   MÉTHODE UTILISÉE : RAKING RATIO
*   LE CALAGE A ÉTÉ RÉALISÉ EN 7 ITÉRATIONS
*   LES POIDS ONT ÉTÉ STOCKÉS DANS LA VARIABLE WGEN DE LA TABLE POIDSGEN
*   ET DE LA TABLE POIDSGEN_INDIV
```